

Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity

Kisung You^{a,1}, Hae-Jeong Park^{b,c,d,e,1,*}

^a Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN, USA

^b Department of Nuclear Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

^c BK21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, Republic of Korea

^d Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea

^e Center for Systems Brain Sciences, Institute of Human Complexity and Systems Science, Yonsei University, Seoul, 50-1, Yonsei-ro, Sinchon-dong, Seodaemun-gu, Seoul 03722 Republic of Korea

ARTICLE INFO

Keywords:

Functional connectivity
Riemannian manifold
Symmetric positive definite
Principal geodesic analysis

ABSTRACT

Common representations of functional networks of resting state fMRI time series, including covariance, precision, and cross-correlation matrices, belong to the family of symmetric positive definite (SPD) matrices forming a special mathematical structure called Riemannian manifold. Due to its geometric properties, the analysis and operation of functional connectivity matrices may well be performed on the Riemannian manifold of the SPD space. Analysis of functional networks on the SPD space takes account of all the pairwise interactions (edges) as a whole, which differs from the conventional rationale of considering edges as independent from each other. Despite its geometric characteristics, only a few studies have been conducted for functional network analysis on the SPD manifold and inference methods specialized for connectivity analysis on the SPD manifold are rarely found. The current study aims to show the significance of connectivity analysis on the SPD space and introduce inference algorithms on the SPD manifold, such as regression analysis of functional networks in association with behaviors, principal geodesic analysis, clustering, state transition analysis of dynamic functional networks and statistical tests for network equality on the SPD manifold. We applied the proposed methods to both simulated data and experimental resting state fMRI data from the human connectome project and argue the importance of analyzing functional networks under the SPD geometry. All the algorithms for numerical operations and inferences on the SPD manifold are implemented as a MATLAB library, called SPDtoolbox, for public use to expediate functional network analysis on the right geometry.

1. Introduction

The prevailing view on the brain is a highly-organized network, composed of interactions among distributed brain regions. Interactions among distributed brain regions are often characterized by the cross-correlation matrix of spontaneous fluctuations observed in the resting state functional magnetic resonance imaging (rs-fMRI) (Biswal et al., 1995). The correlation matrix - a representation of functional brain network - has served as a basis for diverse fields of brain research. The exemplary areas include exploration of brain diseases using functional connectivity (Drysdale et al., 2017; Jang et al., 2016; Lee et al., 2017; Yahata et al., 2016), behaviorally associated brain networks (Kyeong et al., 2014; Smith et al., 2015), dynamic connectivity analysis (Allen et al., 2014; Calhoun et al., 2014; Chang and Glover, 2010;

Cribben et al., 2012; Handwerker et al., 2012; Hutchison et al., 2013; Jeong et al., 2016; Monti et al., 2014; Preti et al., 2017), identification of individuals (Finn et al., 2015), inter-individual variability (Jang et al., 2017) and biomarkers in machine learning (Dosenbach et al., 2010; Drysdale et al., 2017; Finn et al., 2015; Yahata et al., 2016). Partial correlation analysis (inverse covariance matrix) has also been used (Cho et al., 2017; Lee et al., 2017; Qiu et al., 2015) to exclude pseudo-connectivity (spurious connection due to polysynaptic or modulatory effects). In those studies, researchers have analyzed the correlation matrices edge by edge (element of correlation matrix) by considering each edge independently from other edges (Dosenbach et al., 2010; Siman-Tov et al., 2016) or by considering multiple edges together to explain certain dependent variables (Leonardi et al., 2013; Park et al., 2014), particularly, in a machine learning framework (Drysdale et al., 2017; Yahata et al., 2016). All these directions do not much differ in disre-

* Corresponding author at: Department of Nuclear Medicine, Yonsei University College of Medicine, Yonsei-ro 50-1, Sinchon-dong, Seodaemun-gu, Seoul, 03722, Republic of Korea.

E-mail address: parkhj@yonsei.ac.kr (H.-J. Park).

¹ These authors equally contributed.

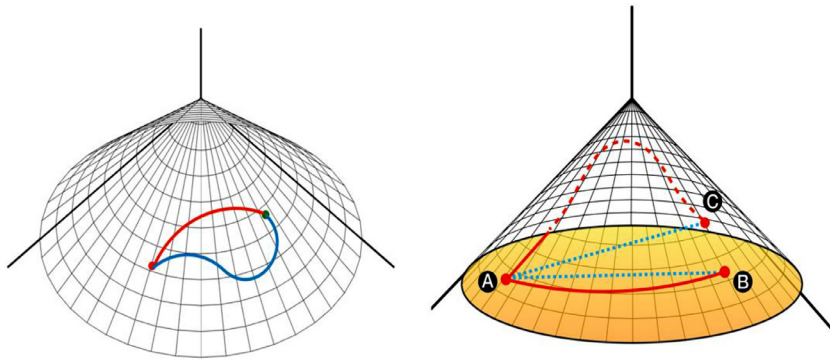


Fig. 1. General SPD manifold. The geodesic distance between A and B is shorter than that of A and C while it is reverse in the Euclidean distance.

garding the topological dependency across elements, a typical property of the correlation matrix.

Mathematically speaking, well-known forms of representation for functional connectivity, such as covariance and precision along with correlation, are symmetric and positive definite (SPD) matrices in common. It is well known that the collection of SPD matrices forms a cone-shape Riemannian manifold (Fig. 1). Accordingly, the operations on the SPD manifold can better be performed based on the corresponding geometric structure rather than the Euclidean geometry. The geometric properties of SPD manifold have not yet been fully recognized in neuroimaging communities. Relatively few but increasing number of studies on the functional connectivity have been conducted on the space of SPD matrices, including evaluating average and variability of group level covariance matrices (Varoquaux et al., 2010; Yamin et al., 2019), statistical testing (Ginestet et al., 2017; Varoquaux et al., 2010), change point detection (Dai et al., 2016), estimation of individual covariance matrix (Rahim et al., 2019), regression of functional connectivity in estimating structural connectivity (Deligianni et al., 2011) and dimension reductions for machine learning (Davoudi et al., 2017; Ng et al., 2014; Qiu et al., 2015; Xie et al., 2017) to name a few.

Despite increasing interest for SPD-based analysis, we still lack access to diverse methods for connectivity analysis on the SPD manifold, such as smoothing and regression, clustering, principal component analysis, state transition analysis of dynamic functional networks and statistical tests. The purpose of the current study is to provide methods for analyzing correlations (or covariance, inverse covariance matrices) on the SPD manifold. We would like to mention one notable characteristic of correlation matrices. Correlation matrix, which is also symmetric and positive definite, is a normalized version of covariance matrix and constitutes a strict subset or submanifold of SPD. To study distinct structure in the constrained set, geometry of ellipsope rather than SPD can be a suitable candidate to study correlation matrices. However, it mainly deals with low-rank structure and has little been investigated (Grubišić and Pietersz, 2007). In this study, we consider the correlation as a constrained type of SPD matrix that forms a submanifold, inheriting geometric properties of general SPD manifold. This does not diminish the importance of SPD manifold and our contribution to study functional connectivity since many representations of functional connectivity are equipped with SPD properties and the proposed methods are presented for a broader class of SPD matrices.

The current paper is composed of four main parts. Firstly, we provide a brief overview and description on the manifold of SPD matrices with a focus on the correlation matrix. Second, we introduce functions for operations or learning algorithms for correlation matrices on the SPD manifold in consideration of practical usages in the functional network analysis. The basic operation starts with definition of geodesic distance between SPD matrices, which bases evaluation of group mean, variation, and statistical tests on the SPD manifold. We also introduce methods for smoothing and regression, principal geodesic analysis and clustering. Third, we present the need to functional connectivity matrices on the SPD space by simulation and by experimental data. All the algorithms are implemented in terms of MATLAB (Mathworks, inc. USA) toolbox,

called SPDtoolbox, which is freely available on GitHub (<https://github.com/kyoustat/papers>) for public use to expediate functional network analysis on the right geometry.

2. Methods

2.1. Review of Riemannian geometry

2.1.1. Elements of Riemannian manifold

In brief, a *manifold* \mathcal{M} is a topological space that locally resembles Euclidean space at each point. Manifold is called *smooth* when transition maps are smooth. A Riemannian manifold (\mathcal{M}, g) is a smooth manifold whose inner product g_p on the tangent space for a point p varies smoothly. From a practitioner's perspective, it is important to keep in mind that Riemannian manifold is not a vector space. That means, we are endowed with a concept of distance that satisfies metric properties on the space that is locally Euclidean-like but it is not Euclidean space.

For example, consider operations on a sphere $S^2 \subset \mathbb{R}^3$ and take north and south poles with coordinates $(0, 0, 1)$ and $(0, 0, -1)$. If we simply treat data residing on the surface with Euclidean structure, averaging coordinates of north and south poles would return $(0,0,0)$, which corresponds to the inner core or origin of the sphere. However, if we want mathematical operations to abide by the spherical geometry, simply 'adding' coordinates of two poles would be an invalid concept. We introduce some concepts and definitions without rigorous mathematical exposition in the following.

A tangent space $T_p\mathcal{M}$ at a point $p \in \mathcal{M}$ is a set of tangent vectors (Fig. 2), which are derivatives of curves crossing p . The statement for a manifold being locally similar to Euclidean space generally refers to the tangent space and its properties as vector space. Two important mathematical operations are exponential and logarithm maps. Let $v \in T_x\mathcal{M}$ be a tangent vector at x . Exponential map $\exp_x(v) \in \mathcal{M}$ defines a unique shortest curve (geodesic) from a point x in the direction of v in \mathcal{M} so that the operation results in a point in \mathcal{M} (Fig. 2). Logarithm map is an inverse of exponential map in that for $x, y \in \mathcal{M}$, $\log_x(y) \in T_x\mathcal{M}$ which corresponds to y when sent back via exponential map. Finally, geodesic distance is the length of the shortest curve that connects two points x and y .

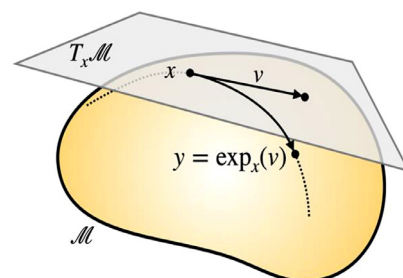


Fig. 2. Exponential map at x in the direction of v on \mathcal{M} .

2.1.2. Fréchet mean and variation

Given the data $x_1, x_2, \dots, x_n \in \mathcal{M}$, the sample Fréchet mean μ_n is defined as a minimizer of the sum of squared distances,

$$\mu_n = \operatorname{argmin}_{p \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \rho^2(p, x_i)$$

where $\rho : \mathcal{M} \times \mathcal{M} \rightarrow (\mathbb{R})$ is a geodesic distance for $x, y \in \mathcal{M}$, the length of the shortest curve on M connecting two points. If Fréchet mean exists, the Fréchet variation V_n is the attained minimum value for μ_n ,

$$V_n = \frac{1}{n} \sum_{i=1}^n \rho^2(\mu_n, x_i)$$

2.1.3. Riemannian metric and geodesic distance

Riemannian metric g is a symmetric (0,2)-tensor field that is positive-definite at every point. Formally speaking, we have

$$g_p : T_p \mathcal{M} \times T_p \mathcal{M} \rightarrow (\mathbb{R}), p \in \mathcal{M}$$

such that for every differential vector fields X, Y on M , $p \mapsto g_p(X|_p, Y|_p)$ is a smooth function and $g(X, X)|_p := g_p(X|_p, X|_p) >= 0$ and $g(X, X)|_p = 0$ if and only if $X|_p = 0$. Locally at a point p , it endows an inner product structure on tangent space. One observation (Pennec, 2006) is that the geodesic distance is closely related to the Riemannian metric in that for $x, y \in M$,

$$\rho(x, y) = \min_{\gamma: [0,1] \rightarrow \mathcal{M}} \int_0^1 \dot{\gamma}(t) dt = \| \log_{g_x}(y) \| = \sqrt{g_x(\log_x(y), \log_x(y))}$$

for γ is a curve minimizing the distance and logarithmic map $\log_p(\cdot) : \mathcal{M} \rightarrow T_p \mathcal{M}$.

2.2. SPD manifold

Our special interest is the space of SPD to which most representations of functional connectivity belong. Conventional measures such as covariance, precision, and correlation matrices all share identical properties.

Definition. $SPD(n)$ is a space of symmetric and positive definite matrices $SPD(n) = \{X \in \mathbb{R}^{n \times n} | X = X^T, \lambda_{\min}(X) > 0\}$ where λ_{\min} denotes the minimum eigenvalue.

It is well known that SPD is, in fact, a Riemannian manifold. That means, there exists certain geometric structures not necessarily similar to that of Euclidean space. In the following sections, we provide mathematical characterization and computational issues from pedagogical perspective.

2.2.1. Matrix exponential and logarithm

For a square matrix X , the matrix exponential is defined as

$$\operatorname{Exp}(X) := \sum_{k=0}^{\infty} \frac{1}{k!} X^k$$

analogous to a real-number case. Matrix (principal) logarithm can also be defined using the definition above by

$$\operatorname{Log}(X) := Y \text{ such that } \operatorname{Exp}(Y) = X.$$

For more details of matrix exponential, logarithm, and their roles in what is known as Lie group in connection with SPD manifold, we refer to Moakher (2005) for interested readers.

Due to its pervasive nature, the space of SPD matrices has been extensively studied (Bhatia, 2015). Many attempts to formalize the non-Euclidean geometry on SPD have been made where distance between two SPD objects are defined by Cholesky decomposition (Dryden et al., 2009), symmetrized Kullback-Leibler divergence (Wang and Vemuri, 2005), Jensen-Bregman LogDet divergence (Cherian et al., 2013) and Wasserstein distance (Takatsu, 2011) to name a few. Among many frameworks, the most popular metrics on the SPD geometric space are the affine-invariant Riemannian metric (AIRM) and the log-Euclidean Riemannian metric (LERM).

2.2.2. Intrinsic geometry with AIRM

Affine-invariant Riemannian metric (AIRM) is one of the most popular geometric structures on the space of SPD matrices (Pennec et al., 2006). As its name suggests, the specific metric, which takes two elements in a tangent space to return a number, characterizes the corresponding geometry of the space. For SPD equipped with AIRM, given two tangent vectors v, w at a point $p \in \mathcal{M} = SPD$,

$$g_p(v, w) = \langle v, w \rangle_p := \operatorname{Tr} (p^{-1} v p^{-1} w)$$

where p^{-1} is standard matrix inverse and Tr a trace operator.

Two important operations that connect the manifold \mathcal{M} and the tangent plane $T_x \mathcal{M}$ at a point $x \in \mathcal{M}$ are exponential and logarithmic maps for which SPD manifold is endowed analytical formula. For a tangent vector v at x , its exponential map $\operatorname{Exp}_x : T_x \mathcal{M} \rightarrow \mathcal{M}$ is defined as

$$\operatorname{Exp}_x(v) = x^{1/2} \operatorname{Exp}(x^{-1/2} v x^{-1/2}) x^{1/2}$$

and logarithmic map at x , $\operatorname{Log}_x : \mathcal{M} \rightarrow T_x \mathcal{M}$ for a point $y \in \mathcal{M}$ is

$$\operatorname{Log}_x(y) = x^{1/2} \operatorname{Log}(x^{-1/2} y x^{-1/2}) x^{1/2}$$

so that the geodesic distance $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$ for two points $x, y \in \mathcal{M} = SPD$ can be computed as

$$d^2(x, y) = g_x(\operatorname{Log}_x(y), \operatorname{Log}_x(y)) = \| \operatorname{Log}(x^{-1} y) \|_F^2.$$

2.2.3. Extrinsic geometry under LERM

The log-Euclidean Riemannian metric (LERM) is another popular machinery along with AIRM (Arsigny et al., 2007). To briefly introduce, LERM framework can be considered as an analysis procedure by embedding points in SPD manifold to Euclidean space via matrix logarithm. For example, geodesic distance for two points $x, y \in \mathcal{M} = SPD$ under LERM is defined as

$$d^2(x, y) = \| \operatorname{Log}(x) - \operatorname{Log}(y) \|_F^2,$$

with $\| \cdot \|_F$ a Frobenius norm. As we noted before, LERM accounts for Euclidean geometry after the matrix logarithm at the identity I . We can check this statement by the fact that $\operatorname{Log}(x) = \log_I(x)$ with conventional inner product in the Euclidean space.

We close this section by making a note that AIRM and LERM are two principled approaches known as *intrinsic* and *extrinsic* analysis of manifold-valued data. While intrinsic approach is strictly bound by the innate geometric conditions and makes all procedures more persuasive, extrinsic approach depends on an embedding to Euclidean space that preserves considerable amount of geometric information, which eases computation and theoretical analysis thereafter. The dichotomy is common in the studies of statistics on manifolds and we suggest (Bhattacharya and Bhattacharya, 2012) to interested readers for detailed description of the dichotomy of principled approaches.

2.3. Algorithms on the SPD manifold

In this section, several learning and inference functions are described with algorithmic details. Across all descriptions in this section, we use the following notations; $\{X_i\}_{i=1}^n \in \mathcal{M}$ for data residing on the SPD manifold, μ and μ_j means of all data and class j , S_i an index set of data who belong to the cluster class i , superscript (t) an iteration t , and $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$ a geodesic distance function as defined in the previous section. Algorithms introduced in this section are largely governed by the intrinsic geometry of AIRM framework. For extrinsic log-Euclidean geometry, most of standard algorithms for data in Euclidean space can be directly applicable after taking matrix logarithmic transformation on all data points as long as the final output matrix in the Euclidean space is taken back via matrix exponential.

2.3.1. Fréchet mean and variation (spd_mean.m)

Given data points $x_i \in \mathcal{M}, i = 1, \dots, n$, empirical intrinsic Fréchet mean μ_n is a minimizer of average squared geodesic distances

$$\mu_n = \operatorname{argmin}_{p \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n d^2(p, x_i) =: f(p)$$

and corresponding variation is

$$V_n = \sum_{i=1}^n d^2(\mu_n, x_i)/n.$$

One of standard methods to obtain μ_n is to use a Riemannian gradient descent algorithm proposed in Pennec et al. (2006), in which conditions for existence and uniqueness are discussed as well as the algorithm. At each iteration, the gradient of f is computed in an ambient space $\nabla f(p) = \sum_{i=1}^n 2\log_p(x_i)/n$ and then projected back onto the manifold by exponential map so that we repeat following steps at iteration t ,

$$W^{(t)} = \frac{2}{n} \sum_{i=1}^n \log_{\mu^{(t)}}(x_i)$$

$$\mu^{(t+1)} = \exp_{\mu^{(t)}}(-W^{(t)})$$

until convergence and the obtained mean μ_n is plugged in to obtain the corresponding variation of our data.

On the other hand, computation of extrinsic Fréchet mean and variation under log-Euclidean geometry is rather straightforward. As mentioned before, under LERM framework, data are embedded into some Euclidean space by matrix logarithmic map in that we have the following one-linear expression to compute objects of our interests,

$$\mu_n = \text{Exp}\left(\frac{1}{n} \sum_{i=1}^n \text{Log}(x_i)\right),$$

$$V_n = \frac{1}{n} \sum_{i=1}^n \|\text{Log}(\mu_n) - \text{Log}(x_i)\|^2$$

where $\|\cdot\|$ is a conventional Euclidean norm.

2.3.2. Test for equality of two distributions (spd_eqtest.m)

It is often a prerequisite to check whether two groups of objects are identically distributed for further analysis. In statistics, this is usually conveyed with hypothesis testing procedures to determine if there exists difference in certain measures of interests, including mean, variance, and so on. One of general test methods is a two-sample testing for the equality of two distributions. When given two sets of independent observations $x_1, \dots, x_m \sim F$ and $y_1, \dots, y_n \sim G$, we test the null hypothesis $H_0 : F = G$ against the alternative $H_1 : F \neq G$.

Among many classes of distribution tests developed for Euclidean data, we employ one nonparametric method by Biswas and Ghosh (2014). This method is solely based on inter-point distances between data points for inference. Maa et al. (1996) showed that testing $F = G$ is equivalent to test whether three distributions D_{FF}, D_{FG}, D_{GG} are identical, where D_{FF}, D_{FG}, D_{GG} denote the distributions of $\|X - X_*\|, \|X - Y\|$, and $\|Y - Y_*\|$ with X_* and Y_* being identically distributed as X and Y . The proof from Maa et al. (1996) is directly applicable to Riemannian manifolds embedded in Euclidean space since the limit in volume form coincides with Euclidean ball locally. This leads to a convenient adaptation of the method by simply replacing the Euclidean norm with geodesic distance on SPD. To briefly introduce the adaptation of aforementioned method, define

$$\hat{\mu}_{D_F} = \left[\frac{2}{m(m-1)} \sum_{i=1}^m \sum_{j=i+1}^m d(x_i, x_j), \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n d(x_i, y_j) \right]$$

$$\hat{\mu}_{D_G} = \left[\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n d(x_i, y_j), \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n d(y_i, y_j) \right]$$

and reject the null hypothesis $H_0 : F = G$ for larger values of the test statistic $T_{m,n} = \|\hat{\mu}_{D_F} - \hat{\mu}_{D_G}\|_2^2$. Since analytical expressions for the sampling distribution of $T_{m,n}$ in the limiting sense is not available, we use the permutation test by rearranging labels of the given data in order to achieve a suitable threshold defining ‘largeness’ of the statistic (Pitman, 1937). In computational aspect, once we have a pre-computed distance matrix $D_{FG} \in \mathbb{R}^{(m+n) \times (m+n)}$ for all observations, permutation testing procedure described above does not demand intensive computing resources since acquiring $T_{m,n}$ consists of scalar addition and multiplication only.

2.3.3. Pairwise coefficient test of equal distribution (spd_eqtestlem.m)

In functional connectivity analysis, it is often of interest to locate which connections show group-level difference at its significance. One common approach is to perform univariate two-sample testing for each entry and apply multiple testing correction. Kim et al. (2014) compared performance of some statistical hypothesis tests in detecting group differences in brain functional networks. The aforementioned approach to employ multiple univariate tests is based on an implicit assumption that each test for pairwise coefficients is independent of each other. Furthermore, asymptotic-based tests depend upon additional assumptions including underlying distribution in that absence of sufficient validation necessitates the use of nonparametric procedure (Wasserman, 2006).

In this study, we introduce a novel procedure for comparing pairwise coefficients under geometric constraints imposed by SPD manifold. Hutchison et al. (2010) proposed a bootstrap procedure for comparing coefficients at the vicinity of Fréchet means by logarithmic mapping to assure independence of parameters. Instead of local comparison, we take an alternative approach to use group-level information represented by Fréchet mean and compare elements thereafter using permutation testing framework. Given data of two groups x_1, \dots, x_m and y_1, \dots, y_n all on $SPD(p)$, the algorithm goes as follows.

- 1 compute Fréchet means μ_x and μ_y , and take $D_{xy} = |\mu_x - \mu_y| \in (\mathbb{R})^{p \times p}$.
- 2 For $t = 1, \dots, T$ iterations,
 - (a) select $\tilde{x}_1, \dots, \tilde{x}_m$ and $\tilde{y}_1, \dots, \tilde{y}_n$ from the data without replacement.
 - (b) compute Fréchet means $\mu_{\tilde{x}}, \mu_{\tilde{y}}$ for permuted samples
 - (c) record $D_{xy}^{(t)} = |\mu_{\tilde{x}} - \mu_{\tilde{y}}|$
- 3 for each (i, j) pair, compute p -values $p(i, j)$ by

$$p(i, j) = \frac{\#\{z \in D_{xy}^{1:T}(i, j) | z \geq D_{xy}(i, j)\} + 1}{T + 1}$$

where $D_{xy}^{1:T}(i, j) = [D_{xy}^1(i, j), \dots, D_{xy}^T(i, j)]$.

2.3.4. k-means clustering (spd_kmeans.m)

k-means algorithm (MacQueen, 1967) is one of famous clustering algorithms for data analysis. As pointed out in Goh and Vidal (2008), the method is easily extensible to non-Euclidean data since it solely depends on the distance measure in determining class memberships. We implemented standard Lloyd’s algorithm (Lloyd, 1982) modified to our context.

- 1 randomly choose k data points as cluster means $\mu_1^{(1)}, \dots, \mu_k^{(1)}$.
- 2 repeat following steps until convergence:
 - assign each observation to the cluster by smallest distances to cluster centers,

$$S_i^{(t)} = \{X_j | d(X_j, \mu_i^{(t)}) \leq d(X_j, \mu_j^{(t)}) \text{ for all } j, 1 \leq j \leq k\}$$
 and when it comes to a situation where an observation can belong to one of multiple clusters, assign the cluster randomly.
 - update cluster centroids by Fréchet means of each class,

$$\mu_i^{(t+1)} = \underset{p \in M}{\text{argmin}} \sum_{j \in S_i^{(t)}} d^2(p, X_j) \text{ for } i = 1, \dots, k$$

where $|S_i|$ refers to the number of observations in class i .

To guide the selection of cluster numbers k , we implement an adaptation of celebrated Silhouette method (Rousseeuw, 1987) as a cluster quality score metric since it only relies on any distance metric (spd_clustval.m). Silhouette is a measure of proximity for observations in a cluster to the points in its neighboring clusters. Each observation is measured with a score between $[-1, 1]$ where large positive number indicates an observation is well separated by the current clustering. For each observation i , a silhouette value is

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where

$$a(i) = \frac{1}{|S_i| - 1} \sum_{j \in S_i, i \neq j} d(X_i, X_j)$$

$$b(i) = \min_{i \neq j} \frac{1}{|S_j|} \sum_{j \in S_j} d(X_i, X_j)$$

From the expression above, $a(i)$ is the average distance between an observation i and the rest in the same cluster of X_i , while $b(i)$ measures the smallest average distance from i -th observation to all points in all other clusters. When a cluster is singleton consisting of a single observation, set $s(i) = 0$. Once silhouette values are computed for all data points, global score is computed by averaging individual silhouette values and a partition of the largest average score is said to be superior to others.

2.3.5. Modularity-based clustering

It should be noted that every execution of k -means clustering shows variations in its result due to initialization. To make a consistent result for multiple k -means clustering and thus make clustering robust to initial values, we adopt modularity analysis of multiple k -means clustering, which was applied to sort neural time courses in our previous study (Jung et al., 2019). The algorithm repeats k -means clustering N times for sample covariance matrices with an empirical K or evaluated by Silhouette score. Then, it evaluates the frequency for each pair of covariances being a same cluster among N times of k -means clustering. This composes a frequency adjacency matrix for every pairs of covariance matrices. The element of the frequency adjacency matrix designates how frequently two nodes (covariances) are clustered as a same cluster. The *modularity optimization* algorithm (Newman and Girvan, 2004), implemented in the Brain Connectivity Toolbox (Rubinov and Sporns, 2010), identifies optimal clusters from the frequency adjacency matrix while maximizing the modularity index Q . The modularity index Q for an adjacency matrix A is defined as:

$$Q = \frac{1}{V} \sum_{ij} \left(A_{ij} - \frac{s_i s_j}{V} \right) \delta_{M_i M_j}$$

where A_{ij} is the edge weight between the i -th and j -th nodes, V is the total sum of weights. $\delta_{M_i M_j}$ is a kronecker delta function that assigns 1 if the i -th and j -th nodes are in the same module ($M_i = M_j$) and 0 otherwise.

In addition to k -means clustering, the SPDtoolbox provides clustering of sample covariance matrices by evaluating an adjacency matrix of negative geometric distances among the samples. The modularity optimization algorithm finds clusters that maximize the modularity index.

2.3.6. Smoothing and regression (spd_smooth.m)

Suppose we have paired data $(t_i, X_i), i = 1, \dots, n$ where $t_i \in \mathbb{R}$ represents one-dimensional covariate such as age or time and $X_i \in \mathcal{M}$ manifold-valued data. One of the most fundamental questions would be to model the relationship between t and X , i.e., $X_i = F(t_i) + \epsilon_i$. This regression framework, however, is not easily defined even for parametric family of $F(\cdot)$ since the concept of subtraction on Riemannian manifold may not be straightforward. Nevertheless, it is needless to emphasize the importance of such problem since the learned regression function may be used for prediction, smoothing, and so on.

A fast yet powerful method was recently proposed by Lin et al. (2017) for manifold-valued response $X \in \mathcal{M}$ and vector-valued explanatory variable $t \in \mathbb{R}^p$. We focus on introducing a special case $p = 1$ for simplicity along with its usage as a nonparametric regression framework against one-dimensional covariate. The key idea is to graft kernel regression (Nadaraya, 1964; Watson, 1964) onto log-Euclidean geometric structure.

Let $J : \mathcal{M} \rightarrow \mathbb{R}^N, N \geq \dim(\mathcal{M})$ be an embedding of our SPD manifold \mathcal{M} onto high dimensional Euclidean space (Fig. 3). A convenient choice for J is an equivariant embedding, which is known to preserve substantial amount of geometric structure (Bhattacharya and Bhattacharya, 2012). For SPD manifold, the embedding is $J(X) = \text{Log}(X)$, which is logarithmic mapping of X at identity I ,

$$J(X) = \text{Log}(X) = I^{1/2} \text{Log}(I^{-1/2} X I^{-1/2}) I^{1/2} = \log_J(X)$$

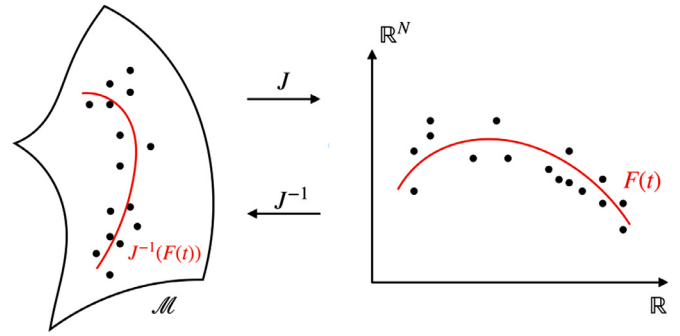


Fig. 3. A schematic description of extrinsic local regression.

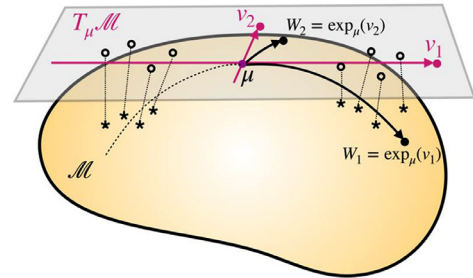


Fig. 4. A schematic description of principal geodesic analysis (PGA).

whose inverse J^{-1} is therefore exponential mapping. Given two mappings, the extrinsic kernel estimate of the regression function F_E is defined as

$$F_E(t) = J^{-1}(\mathcal{P}(F(t)))$$

where

$$F(t) = \frac{\sum_{i=1}^n J(X_i) K_h(t_i - t)}{\sum_{j=1}^n K_h(t_j - t)}$$

for a kernel function $K_h(z) = K(z/h)/h$ with a positive bandwidth $h > 0$ that controls the degree of mass concentration and projection $\mathcal{P} : \mathbb{R}^N \rightarrow J(\mathcal{M})$. Kernel bandwidth h is either fixed or determined via cross-validation. SPDtoolbox provides k -fold cross validation to determine optimal h (spd_smoothcv.m).

2.3.7. Principal geodesic analysis (spd_pga.m)

Principal component analysis (PCA) is a well-known method for exploratory data analysis (Hotelling, 1936; Pearson, 1901). To review in brevity, it aims at finding a basis that captures the largest variance onto which the data is projected. This nature allows us to exploit the method for dimension reduction and feature extraction. Given a data matrix $X = [x_1 | \dots | x_n] \in \mathbb{R}^{p \times n}$ where each row has zero mean, top eigenvectors $w_i, i = 1, \dots, k$, such that $\sum w_i = \lambda_i w_i, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ are taken as basis where $\Sigma = XX^T/n$ is an empirical covariance matrix. The reduced matrix $Y = [y_1 | \dots | y_n] \in \mathbb{R}^{k \times n}$ is obtained by $Y = V^T X$ where $V = [v_1 | \dots | v_k] \in \mathbb{R}^{p \times k}$ is a matrix consisting of eigenvectors as above - also called as principal components or loadings.

Fletcher et al. (2004) grafted the very idea onto the setting where data lies on manifold. Since there is no direct counterpart of empirical covariance matrix and eigen-decomposition on Riemannian manifold, a local approach was first taken. After obtaining a Fréchet mean μ for given data $X_1, \dots, X_n \in \mathcal{M}$, every observation is pulled back onto the tangent space $T_\mu \mathcal{M}$ by $u_i = \text{Log}_\mu(X_i)$ (Fig. 4). Recalling that $T_\mu \mathcal{M}$ is lo-

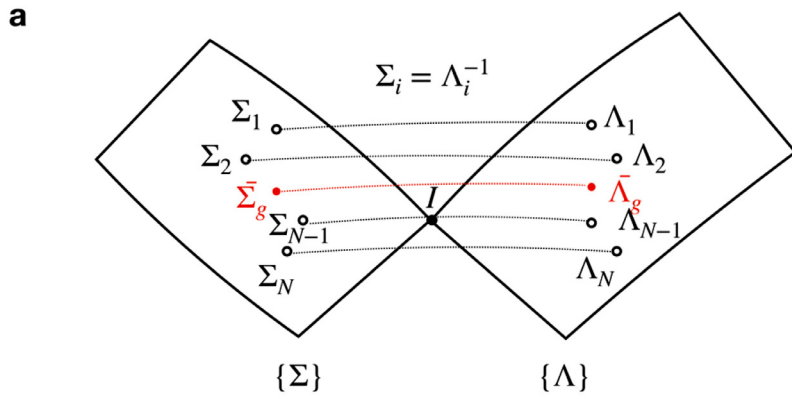
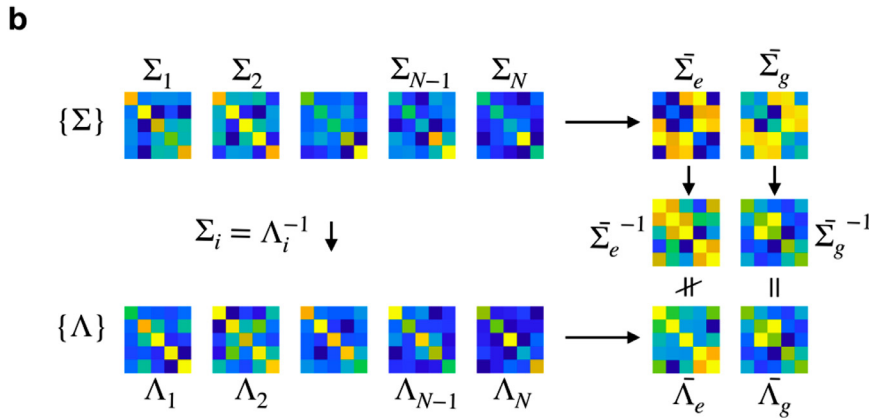


Fig. 5. Concept for covariance and precision (inverse covariance) duality test. The conceptual diagram of duality test (a) and examples (b) for covariance $\{\Sigma_i, i = 1, \dots, N\}$ and precision matrices $\{\Lambda_i, i = 1, \dots, N\}$ are presented. $\bar{\Sigma}_g$ and $\bar{\Lambda}_g$ represent geodesic average and precision matrices. $\bar{\Sigma}_e$ and $\bar{\Lambda}_e$ denote Euclidean average and precision matrices.



cally a vector space, now tangential covariance matrix $S = \sum_{i=1}^n u_i u_i^T / n$ can be obtained whose eigen-decomposition returns $\{\lambda_j, v_j\}$ for $j = 1, \dots, \dim(\mathcal{M})$ where $S v_j = \lambda_j v_j$. This gives local coordinates for each observation $y_i = [v_1 | \dots | v_k]^T \text{Log}_\mu(X_i) \in \mathbb{R}^k$, projected variance λ_j on the tangent space and loadings $W_j = \exp_\mu(v_j) \in \mathcal{M}$ for $j = 1, \dots, k$. The name *principal geodesics* comes from the fact that loadings or principal components are not vectors on $T_\mu \mathcal{M}$ but rather a projected eigenvector by exponential mapping along the geodesic. That means an element y_{ij} from a local coordinate vector $y_i = [y_{i1}, \dots, y_{ik}]$ for an observation X_i induces an approximate weight/significance of i -th observation along j -th geodesic direction. Please note that principal geodesics $\{W_j\}_{j=1}^k$ are not orthogonal to each other as in conventional PCA so that they should be understood as approximate basis in a local sense around the Fréchet mean μ .

3. Results

3.1. Simulation 1. Why SPD geometry should be used?

Other than mathematical characterization of SPD that is clearly different from that of a vector space, we describe why the geometry of SPD endows a natural interpretation with respect to analysis of representations for functional connectivity. Assume we have a population of N covariance connectivity matrices $\{\Sigma\} = \{\Sigma_1, \dots, \Sigma_N\}$. It is of often interest to investigate partial association among different ROIs using precision matrices $\{\Lambda\} = \{\Lambda_1, \dots, \Lambda_N\}$ where each Λ_i is an inverse covariance matrix Σ_i^{-1} . That means we must have $\Sigma_i \Lambda_i = \Lambda_i \Sigma_i = I$ for all i where I is an identity matrix. Then it is natural to question whether averages $\bar{\Sigma}$ and $\bar{\Lambda}$ also satisfy the above relationship $\bar{\Lambda} \bar{\Sigma} = \bar{\Sigma} \bar{\Lambda} = I$ as well. This idea is represented in Fig. 5 that $\bar{\Lambda}$ is an inverse of $\bar{\Sigma}$.

If we consider Euclidean geometry, this is not necessarily true. Considering $\{\Sigma\}$ and $\{\Lambda\}$ as elements in Euclidean space, we can define

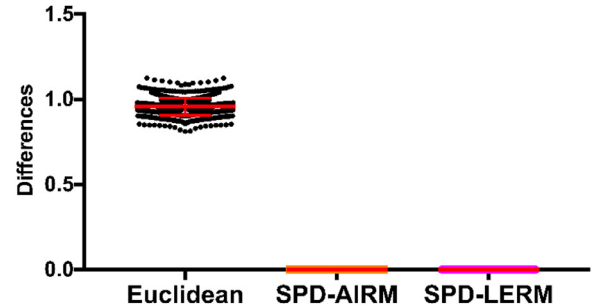


Fig. 6. The result of covariance and precision (inverse covariance) reciprocity test.

averages as follows

$$\bar{\Sigma} = \frac{1}{N} \sum_{i=1}^N \Sigma_i \text{ and } \bar{\Lambda} = \frac{1}{N} \sum_{i=1}^N \Lambda_i = \frac{1}{N} \sum_{i=1}^N \Sigma_i^{-1}$$

in that

$$\bar{\Sigma} \bar{\Lambda} = \frac{1}{N^2} \left\{ \sum_{i=1}^N \Sigma_i \Sigma_i^{-1} + \sum_{i \neq j} \Sigma_i \Sigma_j^{-1} \right\} = \frac{1}{N^2} \left\{ NI + \sum_{i \neq j} \Sigma_i \Sigma_j^{-1} \right\}.$$

Therefore, it is not guaranteed that Euclidean averages of covariance and precision matrices are inverse of each other unless $\Sigma_i \Sigma_j^{-1} = I$ for all $i \neq j$.

On the other hand, averages defined as Fréchet means over $\{\Sigma\}$ and $\{\Lambda\}$ under the geometry induced by AIRM satisfy the above statement under mild conditions. Since matrix inverse is a bijective map while geodesic distance is preserved under inversion, Fréchet mean of precision matrices is an inverse of Fréchet mean of covariance matrices if the mean is unique.

We empirically showed how the choice of geometry and corresponding definition of average/mean leads to the consistency with our intuition. We generated 100 covariance matrices $\Sigma_i, i = 1, \dots, 100$ identically and independently from Wishart distribution $\mathcal{W}_p(V, n)$ with $p = 5$ dimension (network size), $n = 20$ degrees of freedom (the length of time series) for strict positive definiteness, and $V = I_5$ a scale matrix. After computing precision matrices $\Lambda_i = \Sigma_i^{-1}$, we compute both Euclidean and Fréchet means with both AIRM and LERM frameworks on covariance and precision matrices respectively. For each run, the gap is obtained as $\|\bar{\Lambda}\bar{\Sigma} - I_5\|_F$ to see how close $\bar{\Lambda}$ is to $\bar{\Sigma}^{-1}$ where $\|A\|_F = \sqrt{\text{Trace}(A^T A)}$ is a Frobenius norm. If an average precision matrix is truly an inverse of a mean covariance matrix, the gap should be close to 0. Fig. 6 shows distribution of Euclidean and Fréchet gaps on repeated experiments for 1000 times. Empirical observation is that the geometry-induced concept of Fréchet mean aligns with our intuition while simply taking Euclidean averages of covariance and precision matrices does not preserve an inverse relationship of two averages.

3.2. Simulation 2. Modularity-optimization of multiple k-means clustering

As an example of clustering for covariance matrices, we generated three types of covariance matrices with a size of five. We generated 30 covariance matrices from three model matrices C1, C2 and C3 (10 for each). The three covariance matrices (Fig. 7a) were derived from human connectivity matrices obtained from resting state fMRI, which are mutually distant from each other based on all distances under Euclidean, AIRM, and LERM geometries

$$C1 = \begin{bmatrix} 1.0000 & 0.6409 & 0.2196 & 0.1736 & -0.1886 \\ 0.6409 & 1.0000 & 0.7230 & 0.1348 & -0.6583 \\ 0.2196 & 0.7230 & 1.0000 & 0.2734 & -0.6156 \\ 0.1736 & 0.1348 & 0.2734 & 1.0000 & -0.1824 \\ -0.1886 & -0.6583 & -0.6156 & -0.1824 & 1.0000 \end{bmatrix}$$

$$C2 = \begin{bmatrix} 1.0000 & -0.0015 & 0.7898 & 0.7135 & -0.3713 \\ -0.0015 & 1.0000 & 0.2193 & 0.2861 & 0.1240 \\ 0.7898 & 0.2193 & 1.0000 & 0.8958 & -0.3887 \\ 0.7135 & 0.2861 & 0.8958 & 1.0000 & -0.4185 \\ -0.3713 & 0.1240 & -0.3887 & -0.4185 & 1.0000 \end{bmatrix}$$

$$C3 = \begin{bmatrix} 1.0000 & 0.5027 & 0.1763 & 0.2011 & 0.4440 \\ 0.5027 & 1.0000 & 0.0052 & 0.2146 & 0.6810 \\ 0.1763 & 0.0052 & 1.0000 & -0.0690 & -0.0401 \\ 0.2011 & 0.2146 & -0.0690 & 1.0000 & 0.3426 \\ 0.4440 & 0.6810 & -0.0401 & 0.3426 & 1.0000 \end{bmatrix}$$

We then added a zero mean Gaussian noise matrix to each covariance matrix after weighting it by 0.15. The generated covariance matrices are shown in Fig. 7b. For the 30 covariance matrices, we executed 100 k-means clustering with $k = 3$ and constructed a frequency adjacency matrix (Fig. 7c). From the frequency adjacency matrix, modularity optimization identifies three modules and their cluster indices. We extracted cluster centroids for each cluster by taking geodesic average of all members for each cluster (Fig. 7d).

3.3. Simulation 3. Dynamic connectivity and state transition analysis

As an extension of the clustering example, we conducted a state-transition analysis of functional connectivity dynamics (Allen et al., 2014; Damaraju et al., 2014) on the SPD manifold. For a given transition matrix, we simulated a state sequence (length = 30) consisting of three heterogeneous states which correspond to initial covariance matrices, C1, C2 and C3 (Fig. 8a). The transition matrix used is

$$T = \begin{bmatrix} 9 & 2 & 3 \\ 4 & 4 & 1 \\ 2 & 2 & 2 \end{bmatrix}$$

For the 30 correlation matrices (Fig. 8b), we conducted a modularity-optimization of multiple k-means clustering. The method generates a

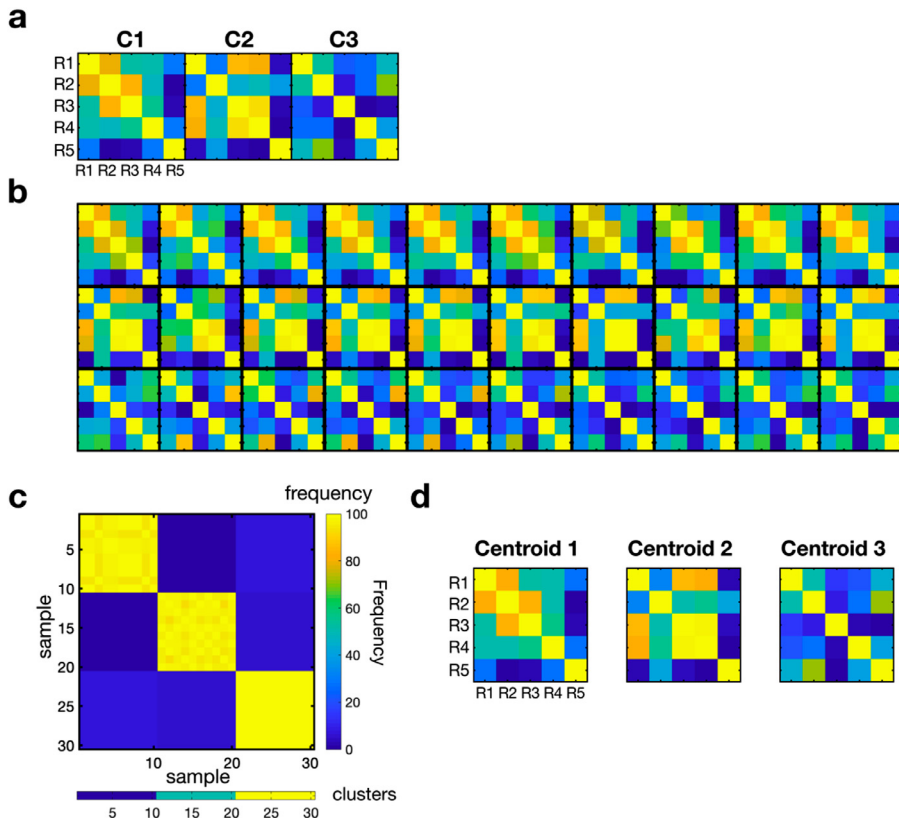


Fig. 7. An example for clustering covariance matrices. Three initial covariance matrices (a) were used to generate 30 random variation of covariance matrices around initial covariance matrices (10 for each C1, C2, and C3) (b). (c) displays the frequency adjacency matrix of the 30 covariance matrices used for modularity optimization. The frequency adjacency matrix indicates the frequency for each pair of covariances being a same cluster after 100 repeated k-means clustering. (d) displays the centroids for the three covariance groups.

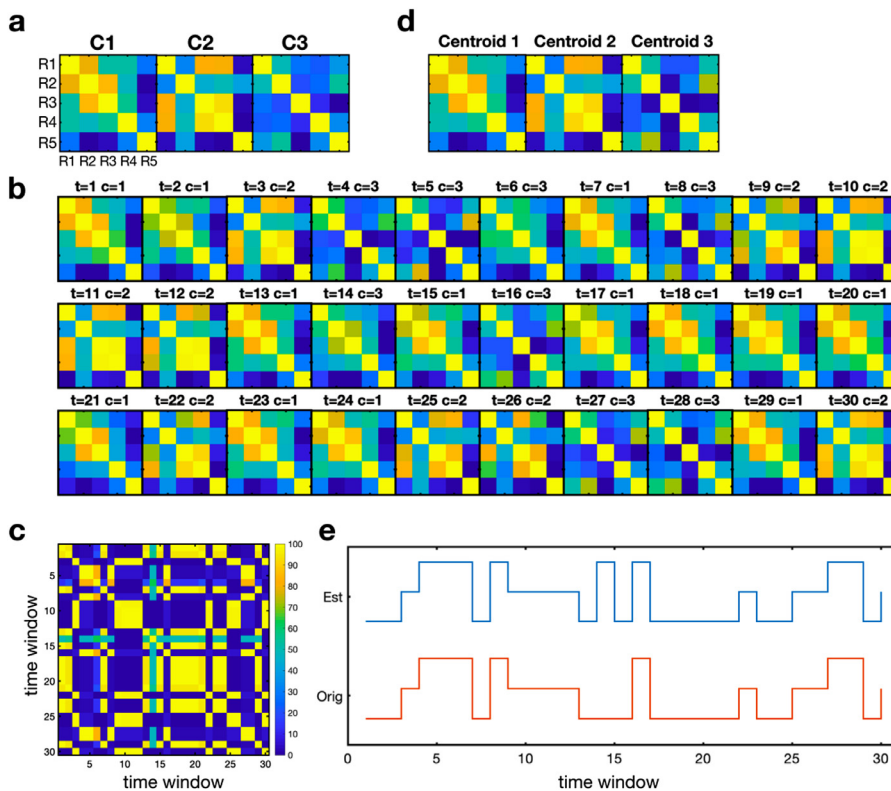


Fig. 8. State transition analysis of dynamic connectivity. Based on three main correlation matrices C1, C2 and C3 (a), total 30 covariance samples were generated according to a state transition sequence (b). The modularity-based k -means clustering was done after composing a frequency adjacency matrix (c) for 100 iterations of k -means clustering. The resulting centroids are displayed in (d) and the sequences of original (red) and estimated (blue) ones are displayed in (e) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

frequency adjacency matrix by applying 100 repetitions of k -means clustering (Fig. 8c), which was used for modularity optimization. The cluster centroids are displayed in Fig. 8d and the initial sequence and resulting sequence are displayed in Fig. 8e.

3.4. Simulation 4. Smoothing and regression

We conducted smoothing of simulated noisy covariance matrices. Covariance matrices simulated were geometrically positioned between the two covariance matrices C1 and C2 with an exponentially increasing distance weight w ($w = t^2$ for $t = [0, 0.1, 0.2, \dots, 1]$) for C2 and $1-w$ for C1 on the tangential plane (Fig. 9a). We then added a zero mean Gaussian noise matrix to each covariance matrix after weighting it by 0.1 and projected the noisy covariance matrices back onto the manifold to make them as perturbed observations along the geodesic. The generated covariances and their noisy matrices are shown in Fig. 9b. We conducted 5-fold cross-validation to determine the kernel bandwidth h from an exponentially equi-spaced interval of length 10 from $\exp(-1)$ to $\exp(1)$. For each level of h , we applied smoothing function to the noisy covariance matrices. The results for smoothing and bandwidth selection are presented in Fig. 9b and 9d.

We utilized this approach to a regression problem (Fig. 10). We squared 50 uniformly distributed weights and used them as interpolation weights w for geometric locations between the two covariance matrices C1 and C2. We then added a Gaussian noise in the tangent space and projected noisy matrices to suffice SPD condition. Using the bandwidth h corresponding to the minimal k -fold cross-validation error, we estimated an interpolation function, which predicts the covariance matrix at any geometric location (corresponding to a weight w).

3.5. Simulation 5. Principal geodesic analysis

We generated 30 covariance samples by interpolating three target covariances (C1, C2 and C3) with random weights (Fig. 11a). The weights were derived from a uniform distribution but the weights for C1 were

multiplied by 1.2 from the initial weights to emphasize variations of C1. The weighted sum was done in the tangential space. The result covariance samples are presented in Fig. 11b. We then applied principal geodesic analysis to the covariance matrices to find principal geodesics and their weights for the samples (Fig. 11c).

3.6. Simulation 6. Test of equal distributions

In statistical literature of hypothesis testing, simulation study is often adopted to demonstrate the power of a proposed procedure (Lehmann and Romano, 2010). Among several indices, we report empirical type 1 error, which measures ratio of false rejections when the null hypothesis of $H_0 : F_x = F_y$ is true. In our simulation, we use two exemplary tests with samples drawn from the correlation matrices C2 and C3. For each group, we generate time series samples from Wishart distribution by using designated correlation matrix as a scale parameter and ν degrees of freedom (time series length) and normalize the matrices to have unit diagonals. We used $\nu = 50$ to guarantee strict positive definiteness. We varied the number of observations (corresponding to subjects in the group study) belonging to each set from $n = 5$ to 50 in order to check an impact of sample size on the performance. For $T = 500$ iterations, two sets of correlation matrices with varying sample sizes from $n = 5$ to 50 are randomly generated in accordance with the procedure described above and p-values are recorded. Then, empirical type 1 error at the pre-defined significance level α is obtained by

$$\text{Type 1 Error} = \frac{\#\{\text{computed } p\text{-values} \leq \alpha\}}{T}$$

and we conclude the test is valid if empirical type 1 error is similar to α . As shown in Fig. 12, both examples show similar levels of type 1 errors and $\alpha = 0.05$ for different numbers of observations.

3.7. Experimental data analysis

For the experimental data analysis, we used 734 rs-fMRI data (2016 release, male: N=328, female: N=405, age: 22~37, mean: 28.7, std:3.7)

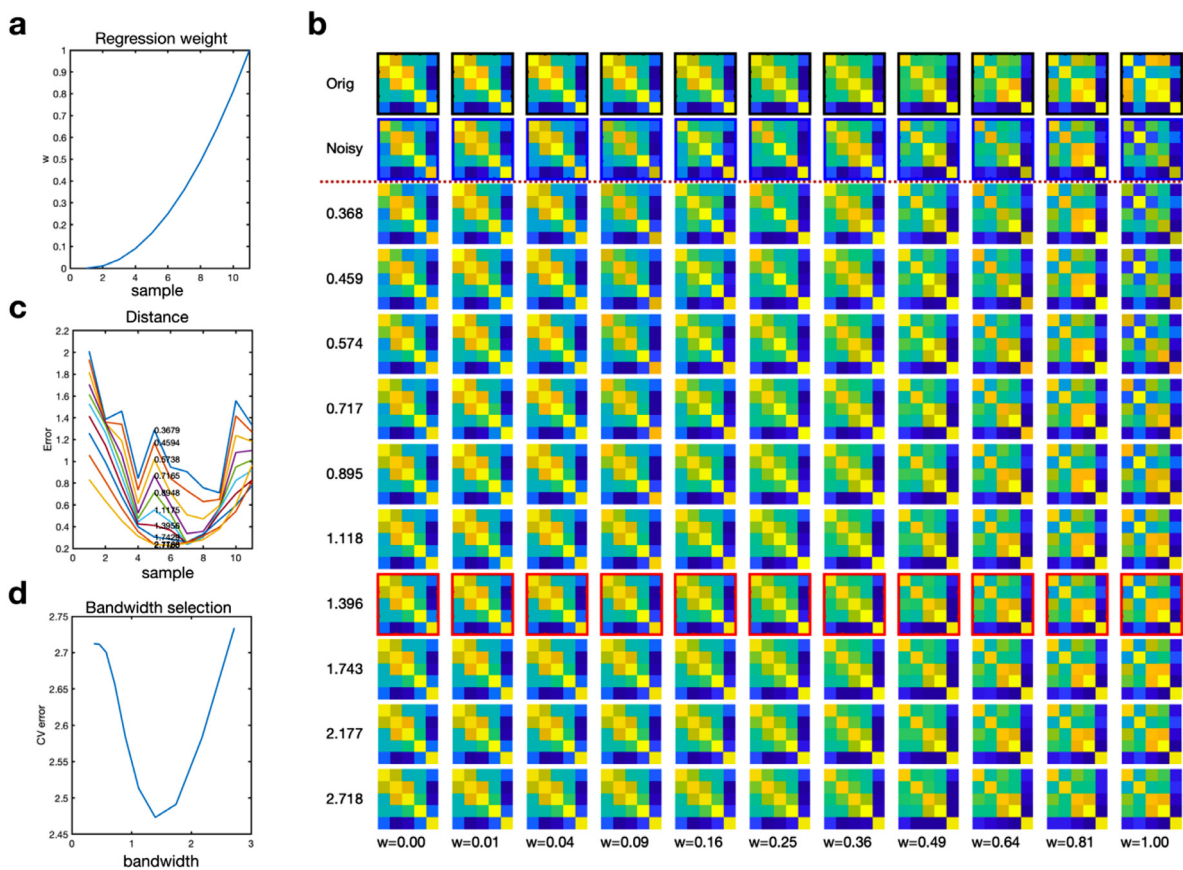


Fig. 9. An example for smoothing of continuously varying correlation matrices with Gaussian noises. A set of noisy continuous covariance matrices (the first row in (b)) were generated by weighted summing of two covariance matrices (C1 and C2 in Fig. 7) with a Gaussian noise in the tangent plane. The weight is presented in (a). For ten different bandwidths, the distances between samples and their corresponding regressed ones are presented in (c). The 5-fold cross validation (CV) errors for all bandwidths are presented in (d). The original covariance samples are displayed in the first row and their noisy versions are displayed in the second row of (b). The lowest CV error was found at the bandwidth of 1.396, the 9th row in (b).

from the Human Connectome Project (HCP) database (Van Essen et al., 2012). All data was sampled at 0.72 Hz during four sessions, with 1200 time points per session. rs-fMRI data was preprocessed according to the HCP minimal preprocessing pipeline and mapped into cortical surfaces (Glasser et al., 2013). For the analysis of brain networks, we extracted time series at ten brain regions in the cortical surface-based atlas (Desikan et al., 2006) corresponding to the default mode network (DMN). They are the superior frontal gyrus (SFG, also known as the dorsomedial prefrontal cortex), medial orbitofrontal cortex (MOF, corresponding to ventromedial prefrontal cortex), parahippocampus (as a part of the hippocampal formation), precuneus and inferior parietal lobe in each hemisphere (Fig. 13A), which are commonly found in previous studies of DMN (Buckner et al., 2008; Power et al., 2011; Yeo et al., 2011). The first eigenvariate derived from the principal component analysis of fMRI time series at surface nodes for each region was used as a regional BOLD signal summary for the region. Pearson correlation of rs-fMRI was used as a measure for functional connectivity of the time series. The geodesic average of all the 734 rs-fMRI data and exemplary functional connectivity (correlation) matrices of 30 subjects are presented in Fig. 13b and 13c.

3.8. Experiment 1. Regression analysis

We conducted a regression analysis of connectivity matrices according to behavior scores. Among many behavior scores in the HCP database, we chose Penn Matrix Test (PMAT) and Pittsburgh Sleep Quality Index (PSQI) to test the feasibility of geodesic regression analysis for SPD. The PMAT24 which was developed by

Gur and colleagues (Bilker et al., 2012) is a shortened version of the Raven's Progressive Matrices that measures fluid intelligence. To represent fluid intelligence, number of correct responses (PMAT24_A_CR) and total skipped items (PMAT24_A_SI) were used. The PSQI test is a widely used 19 self-rated questionnaire that assesses sleep quality and disturbances over one month time interval (Buysse et al., 1989). In the SPD regression analysis, we determined the kernel bandwidth by utilizing 5-fold cross validation at 20 exponentially equidistant bandwidths (Fig. 14a, c and e). With the kernel bandwidth showing minimal CV errors (for concave CV-error curves, e.g., PMAT24-A-CR and PMAT24-A-SI) or knee of CV-error curves (for curves without any local minima, e.g., PSQI-scores), 734 connectivity matrices at different scores for each measure were used to estimate regression function. The representative connectivity matrix at each score according to the learned regression function is presented in Fig. 14b, d and f.

3.9. Experiment 2. Principal geodesic analysis (PGA)

Principal geodesic analysis was applied to the 734 connectivity matrices, from which nine principal geodesic components were derived. The explained variances of all nine components and top three principal geodesic components and their coefficients (the weight of connectivity matrices projected onto each component) are displayed in Fig. 15a–c. The component 3 differs from the other components in that the module composed of the bilateral inferior parietal lobe and the precuneus are differentiated from other brain regions. The component 2 differs from the component 1 in the isolated medial orbitofrontal cortex. The

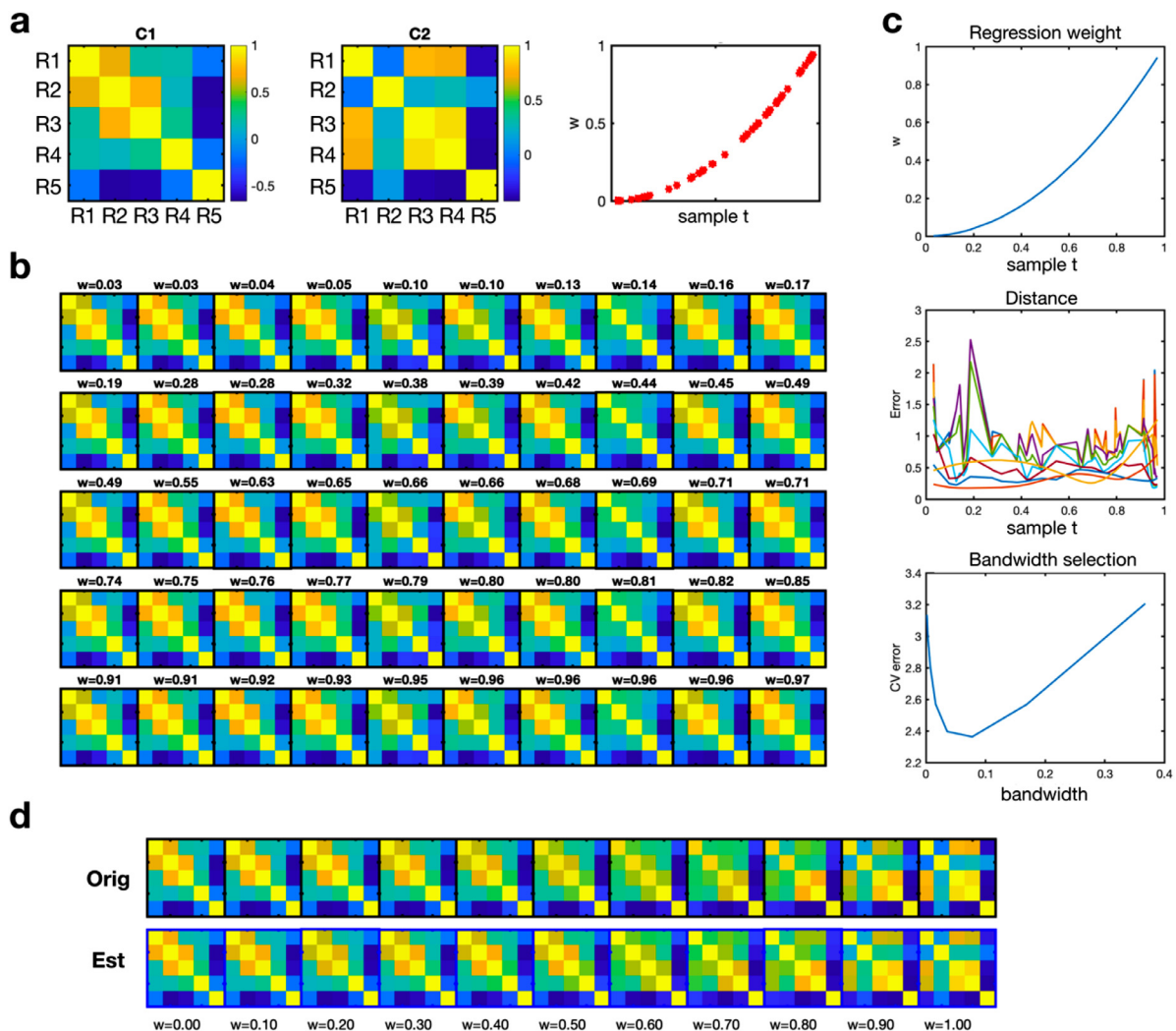


Fig. 10. An example for regression analysis of covariance samples. By weighted summation of C1 and C2 covariance matrices (a) using a square weighted function w (the third panel of a), followed by adding Gaussian noise, we generated 30 samples of correlation matrices (b). The evaluation of regression performance was done at the weight points shown in (c). For ten different bandwidths, the distances between samples and their corresponding regressed ones are presented in the middle panel of (c). The 5-fold cross validation (CV) error for each bandwidth is presented in the bottom panel of (c). The original covariance samples and regressed correlation matrices (with an optimal bandwidth) at ten uniformly spaced weight points are displayed in (d).

geodesic component 1 explains most portions of variability of connectivity matrices as shown in the distribution of coefficients (Fig. 15c). We also evaluated Euclidean principal component analysis for the covariance matrices as a set of vectors (one per each subject) without considering covariance architecture. The results are presented in Fig. 15d–f. The distributions of components and their coefficients in the Euclidean analysis differ from those of geodesic analysis.

3.10. Experiment 3. Clustering

Modularity-optimization of multiple k -means clustering was applied to 734 connectivity matrices. A total of 100 runs of k -means clustering constitutes a frequency adjacency matrix (Fig. 16a), which was sorted by modularity optimization (Fig. 16b). The frequency adjacency matrix indicates how much each pair of nodes is assigned to a same cluster for partitions from k -means clustering. Connectivity matrices of 734 subjects are categorized into three clusters according to this approach. The geodesic mean and Euclidean mean for each cluster and their difference are presented in Fig. 16c–e. The cluster centers 1, 2, and 3 match highly with geodesic principal components 1, 3 and 2 in Fig. 15.

3.11. Experiment 4. Dynamic connectivity

Dynamics in the functional connectivity was evaluated by clustering sliding-windowed connectivity matrices, followed by calculating a transition matrix among the clusters. rs-fMRI time series data (1200 temporal scans) from a subject was divided into 83 epochs by a sliding window approach with a window size of 42 (30 s, 0.72 sec/sample) and an overlap of 28 (20 s) (Fig. 17a). We calculated correlation matrices for all the epochs. Total 100 k -means clustering were applied to the 83 correlation matrices in the SPD space to compose a frequency adjacency matrix (Fig. 17b), which was used for modularity optimization (Fig. 17c). We presented the three geodesic cluster centers (Fig. 17d) and the transition matrix that shows the number of transitions among the clusters (Fig. 17e).

3.12. Experiment 5. Performance comparison between AIRM and LERM

We presented the reliability and computational burden innate to intrinsic Fréchet mean computation using AIRM against log-Euclidean counterpart (LERM) in Fig. 18. It is common practice to take extrinsic

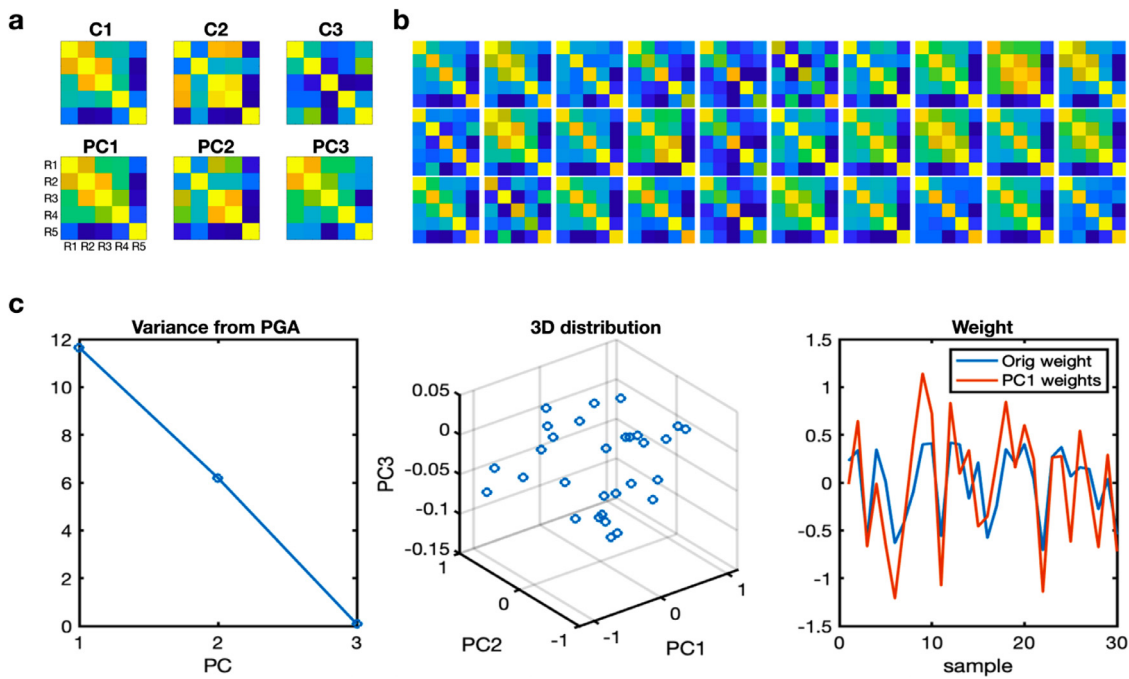


Fig. 11. An example for principal geodesic analysis (PGA) of correlation matrices. 30 correlation matrices in (b) were generated by randomly weighting C1, C2 and C3 correlation matrices in (a). The weights for C1 were multiplied by 1.2. The explained variances of principal components are displayed in the first panel of (c). The sample distributions in the principal component axes are shown in the second panel of (c). The initial weights for C1 and weights for the first principal component (PC1) are displayed in the third panel of (c).

mean as a starting point for gradient descent algorithm due to its asymptotic convergence (Bhattacharya and Bhattacharya, 2012). This means that the strategy we take inevitably requires additional computing time for an iterative scheme but it may be negligibly small by taking a good starting point. We compare the performance of two aforementioned frameworks in computing Fréchet mean of 1000 covariance matrices $\Sigma_i, i = 1, \dots, 1000$ independently generated from Wishart distribution $W_p(V, n)$ for a range of dimensions (network size) $p = 50, 100, \dots, 500$, $V = I_p$ a scale matrix with $n = 10000$ degrees of freedom which is sufficient to guarantee the definiteness of generated samples. The mean squared difference between mean covariances estimated by using AIRM and LERM is presented in the left panel of Fig.18 and the CPU time in seconds are displayed in the right panel of Fig. 18. The experiment was conducted on a personal computer with a quad-core Intel i7 CPU

running at 4 GHz and 24GB RAM (Apple iMac, 2014) using MATLAB R2019b without any explicit parallel computing involved.

4. Discussion

Use of correlation matrix for resting state or task-dependent fMRI has been a basic tool for functional brain network analysis. The implicit assumption of conventional functional network analysis is that each edge is constructed (and can be dealt with) separately from other edges by repetitive pairwise correlation analysis. It has little considered the correlation matrix as a whole and often neglects correlated nature of edges. In some applications, multiple edges are modeled to be jointly involved in the brain diseases or cognitions. However, the vectorization-based multivariate approach does not acknowledge that each edge itself is a

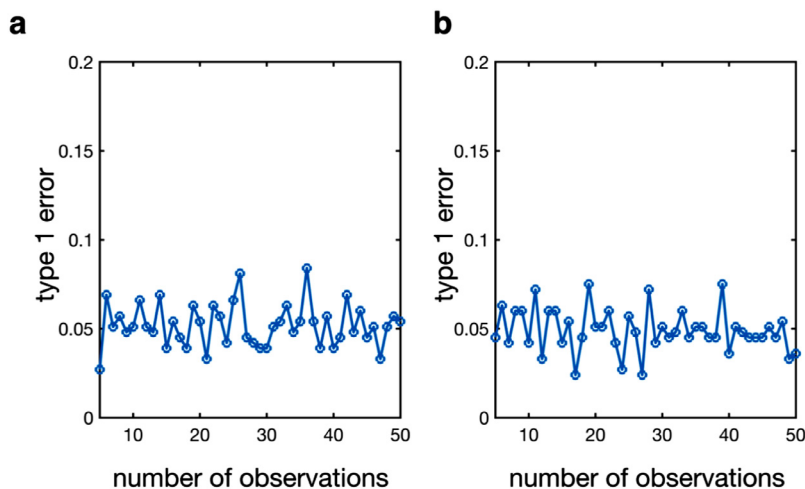


Fig. 12. Test of equal distributions. (a) and (b) show empirical type 1 errors with increasing number of observations in the x-axis for simulations with C2 and C3 at the significance level $\alpha = 0.05$.

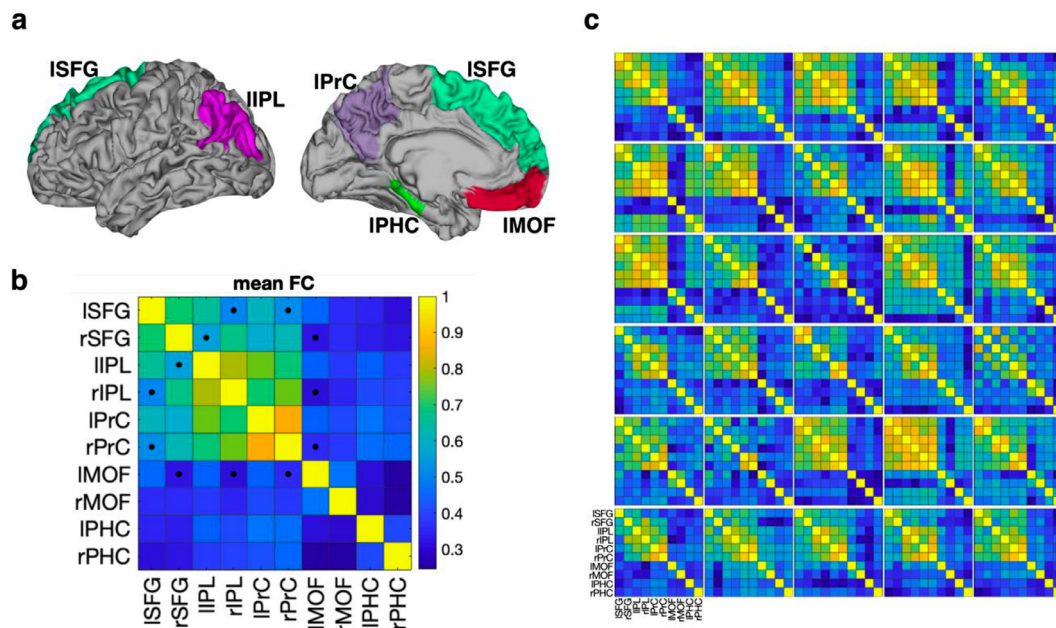


Fig. 13. The brain regions used in the evaluation of the default mode network. a) The brain network includes the superior frontal gyrus (SFG), medial orbitofrontal cortex (MOF), parahippocampus (PHC), precuneus (PrC) and inferior parietal lobe (IPL) according to the Desikan-Killiany atlas (Desikan et al., 2006). The brain regions in the left hemisphere are presented. The first letters “l” and “r” in each region’s name indicate left and right hemisphere. b) The geometric average of a group of 734 functional connectivity matrices (FC) is presented. All square pixels in the connectivity matrix indicate edges among the brain regions. Dots in (b) indicate edges that show 5% differences between Euclidean average and geometric average of group connectivity matrices. c) Functional connectivity matrices of 30 exemplary subjects are displayed.

result of multi-nodal activities and has interrelatedness with others in consequence. It is not difficult to find neurobiological examples that the basic assumption does not hold. For example, some nodes (such as the ventral tegmental area and the substantia nigra pars compacta in the dopamine pathway or the raphe nucleus in the serotonin pathway) play modulatory roles for multiple edges (thus, they are interrelated). Therefore, a single edge may well be considered in association with other edges. This is the rationale to deal with correlation matrix as a whole.

Since the correlation matrix is a type of SPD objects which have a cone-shape Riemannian geometry, we need to take advantage of dealing correlation matrices as manifold-valued objects with corresponding geometric structure. Persistent inversion of covariance matrices for groups in Figs. 5 and 6 is such an example. We can also find several examples on the functional connectivity analysis using the correlation matrix on the SPD manifold (Dai et al., 2016; Dai et al., 2020; Deligianni et al., 2011; Ginestet et al., 2017; Ng et al., 2016; Qiu et al., 2015; Rahim et al., 2019; Varoquaux et al., 2010; Yamin et al., 2019).

Varoquaux et al. (2010) introduced Fréchet mean and variation of correlation matrices to detect functional connectivity difference in the post-stroke patients using group-level covariance averages and demonstrated an increased statistical power in the detection compared to the Euclidean space. They also proposed a bootstrap procedure for pairwise coefficient analysis on the tangent space to ensure mutual independence amid coordinates. In compared to their study, we propose bootstrapping directly on intra-group differences to infer statistical significance of edge difference. Rahim et al. (2019) utilized SPD geometry in estimating functional connectivity of individuals by pulling all covariance matrices to tangent space around Fréchet mean and constructed an empirical prior derived from the tangentialized data to be used as a group-level shrinkage mechanism for individual covariances. In estimating behaviors using covariance matrix, Ng et al. (2016) utilized a whitening transport approach which applies log mapping along the identity matrix after removing the reference covariance effect instead of conventional parallel transport approach. Ginestet et al. (2017) proposed a method to hypothesis testing for binarized connectivity matrix in the SPD space

using the graph Laplacian matrix which inherits SPD structure by its nature. Yamin et al. (2019) compared brain networks between twins with geodesic distance on the SPD manifold. Deligianni et al. (2011) proposed a regression analysis on the SPD manifold in inferring brain functional connectivity from anatomical connections. The change point detection has also been conducted in the SPD space by Dai et al. (2016) who conducted stationarity testing of functional connectivity obtained by the minimal spanning tree in conjunction with resampling-based decision rules. Dai et al. (2020) also proposed a metric for trajectories on the SPD manifold to learn individuals’ dynamical functional connectivity at the population level based on distance-based learning methods. This mainly differs from current smoothing approach proposed by Lin et al., (2017) in that we stand in line for a nonparametric modeling of the smoothed trajectory itself. Besides fMRI study, Sadatnejad et al. (2019) introduced a clustering method on the SPD to represent EEG. Several studies have conducted dimension reduction for machine learning in the SPD space. Qiu et al. (2015) conducted a manifold learning on functional networks using regression upon age after transforming covariance to the Euclidean space by applying logarithmic map at the identity matrix. Davoudi et al. (2017) suggested a dimensionality reduction based on distance preservation to local mean for symmetric positive definite matrices. Xie et al. (2017) also conducted classification after dimension reduction based on bilinear sub-manifold learning of SPD.

The current study can be in line with and an extension of those previous studies, by introducing more algorithms in particular to functional brain network analysis, which were previously used or conceptualized in the SPD research fields. The inference algorithms we newly introduced in the functional connectivity analysis are smoothing and regression, principal geodesic analysis, clustering and statistical tests for correlation-based functional connectivity.

The regression of SPD is to find a relationship between a covariate and covariance matrix, which was generally approximated by minimizing errors (measured data minus predicted ones with a learned regression function). This regression framework for SPD is not trivial since the concept of subtraction may not be straightforward. Lin et al. (2017) pro-

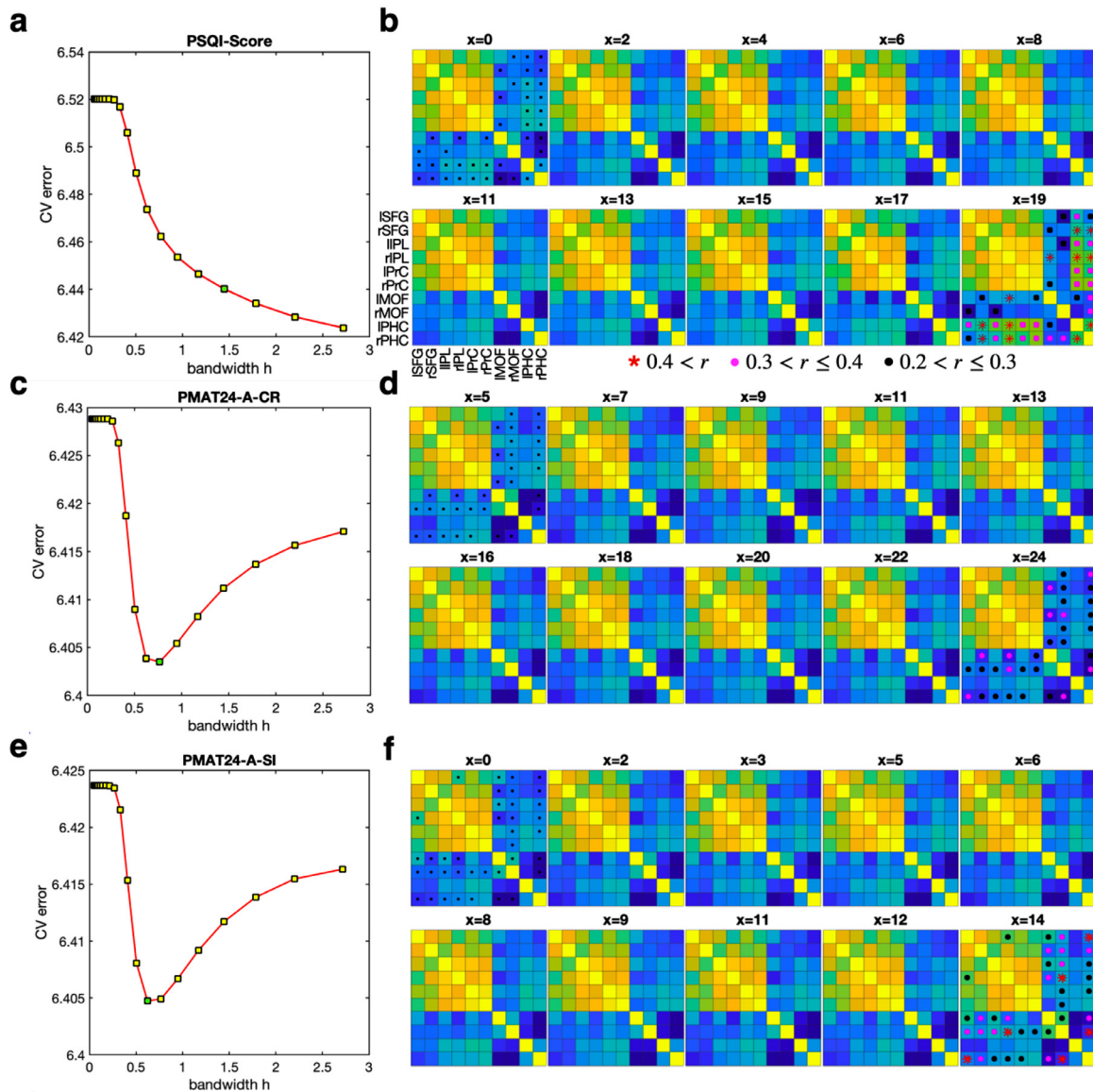


Fig. 14. Geodesic regression analyses for connectivity matrices with Pittsburgh Sleep Quality Index (PSQI) score (a and b), Number of Correct Responses of Penn Progressive Matrices (PMAT24_A_CR) (c and d) and Total Skipped Items of Penn Progressive Matrices (PMAT24_A_SI) (e and f). To optimize the regression kernel bandwidth, 5-folds cross-validation (CV) errors were calculated for every bandwidths h (a, c and e). The bandwidths were set to be exponentially increasing from $\exp(-3)$ to $\exp(1)$ in 20 steps. The bandwidth that have minimal CV errors or curve knee points were chosen to be optimal (green rectangle). The regression results were presented in the form of connectivity matrix for a 10 equispaced scores (x) from the minimum to the maximum scores (b, d and f). To designate changes in the edges for visualization purpose, we marked edges that have 20%, 30% and 40% change ratios between the connectivity matrices of the lowest (the regressed connectivity for the first score, C_1) and highest scores (the regressed connectivity for the last score, C_n). The ratio r was defined by the absolute difference between the connectivity for the first and last scores, normalized by the mean connectivity across all scores, $r = \frac{|C_n - C_1|}{\frac{1}{N} \sum_{i=1}^N C_i} \times 100$, where C_i indicates the representative (regressed) connectivity matrix at the score i (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

posed a powerful regression method associating manifold-valued response and vector-valued explanatory variables by applying kernel regression onto transformed responses under some embedding that preserves geometric information (Nadaraya, 1964; Watson, 1964). We utilized this nonparametric regression framework to associate functional connectivity against one-dimensional covariate. Although we focused on a one-dimensional covariate for simplicity in the current study, this regression can be easily extended to cases with multiple covariates. This approach differs from most of previous studies on regression analysis of functional networks, which have dealt with correlation matrices element (edge) by element (edge) separately from each other edge (Dosenbach et al., 2010; Siman-Tov et al., 2016). The introduced regression also differs from Qiu et al. (2015) which performs Locally Linear

Embedding to embed manifold-valued covariates into Euclidean space so that conventional regression models can directly be used. However, the method by Qiu et al. (2015) does not guarantee straight-forward interpretation of acquired regression coefficients.

In the SPD smoothing and regression analysis, the main issue is how to determine the kernel bandwidth (or smoothness) residing in the SPD data. Since the current geodesic smoothing is a data-driven approach, we opted for a cross-validation scheme to determine the kernel bandwidth ‘ h ’. When we applied the proposed regression analysis to functional networks according to behavior scores, i.e., the PSQI scores or PMAT scores, it estimates meaningful architecture in the functional network. The result suggests the advantage of SPD regression in exploring how functional network changes along a covariate as a whole, not ele-

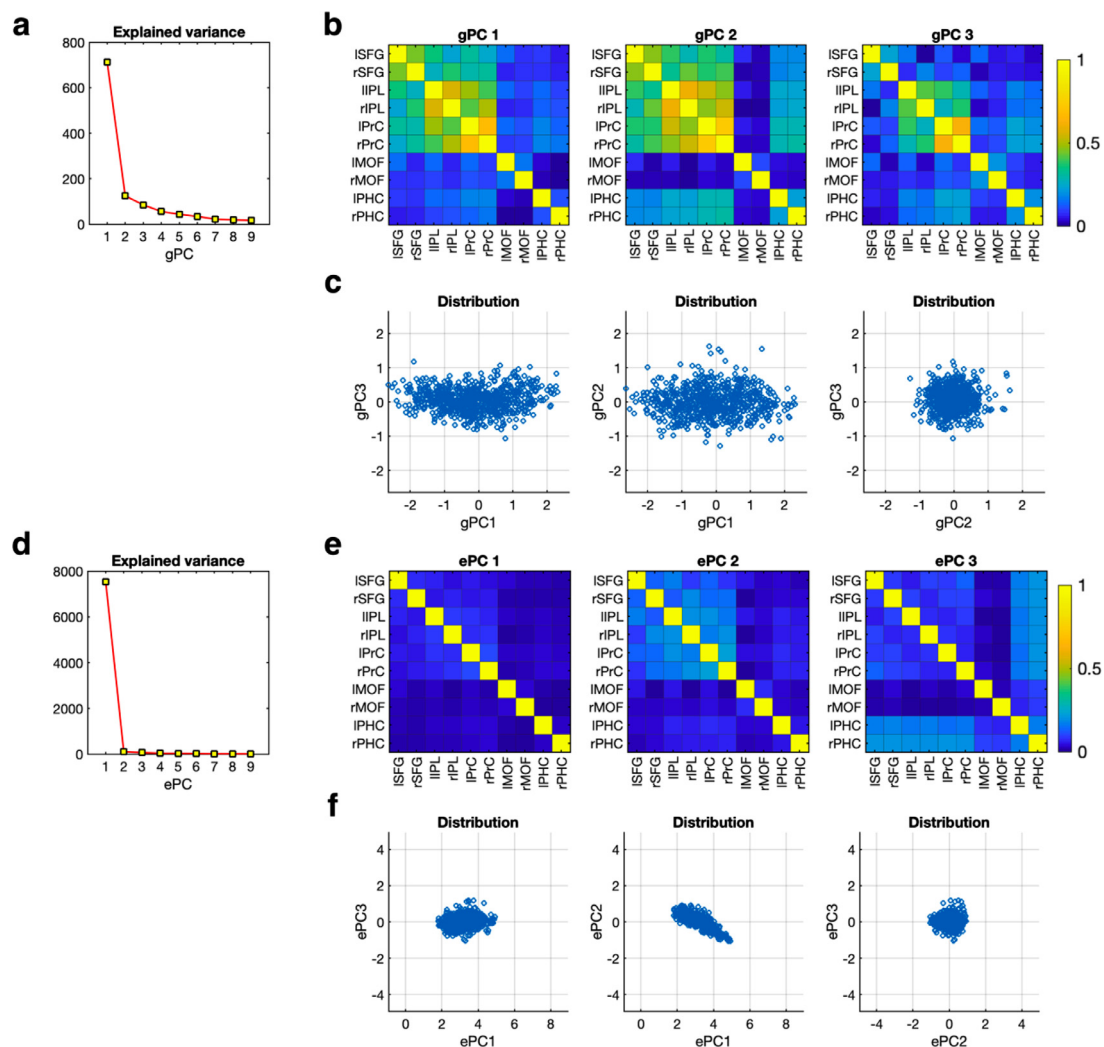


Fig. 15. Principal geodesic analysis and Euclidean principal component analysis of connectivity matrices. The explained variance for 9 components is presented in (a). The top three principal geodesic components (gPC) derived from 734 connectivity matrices are presented in (b). The distributions of coefficients for all connectivity matrices projected onto the top three components are presented in (c). Euclidean principal component analysis shows different explained variance curve (d) and components (e) compared to those of geodesic component analysis. Distributions of 734 coefficients for the three principal components, ePC1, ePC2 and ePC3, are displayed in (f).

ment by element. In the current study, we evaluated the performance of the current smoothing and regression analysis for a prespecified noise level. Further elaboration with a more extensive range of noise levels will be worthy in understanding the performance of the proposed method.

As a multivariate analysis of covariance matrices, principal component analysis (Leonardi et al., 2013) or independent component analysis (Park et al., 2014) has been used in the Euclidean space. Fletcher et al. (2004) introduced a principal geodesic analysis onto the setting with manifold-valued data. Since there is no direct counterpart of empirical covariance matrix and eigen-decomposition on Riemannian manifold, a local approach was first taken in the study of Fletcher et al. (2004). So far, the principal geodesic analysis has been applied to shape analysis (Aljabar et al., 2010; Fletcher et al., 2004) and diffusion tensor analysis (Fletcher and Joshi, 2004). In this study, we introduce principal geodesic analysis to the analysis of functional networks in the form of correlation matrices. As shown in Fig. 15, there are significant differences in the principal components and their coefficients between the two approaches. As a limitation in principal geodesic analysis for the network analysis, not much is known for noise effects (in the covariance matrix) on the principal geodesic analysis. We can-

not disregard the computational cost for the large-sample, large-scale network analysis.

k -means, though powerful in the clustering, is known to suffer from two major drawbacks. First, it is NP-hard in general so that finding the optimal solution is troublesome (Aloise et al., 2009). Lloyd's algorithm provides a heuristic approach to obtain suboptimal solutions given an initialization but outcome variability cannot be removed by different initial partitioning (Celebi et al., 2013). Although many variants have been proposed to overcome such instability including a popular initialization method k -means++ (Arthur and Vassilvitskii, 2007), we still choose random initialization scheme since extra computational complexity by such methods can be high especially in our context. Some methods cannot be directly transferred due to our geometric restrictions where the structure of Euclidean geometry cannot be directly transferable. We can achieve good-enough solutions via parallel computing on multiple initializations. The second drawback of is the selection of cluster numbers k . Unlike probabilistic and Bayesian methods where statistical inference on k is achievable within their modeling framework, a family of k -partitioning algorithms - k -means and its variants - requires post hoc model selection procedures to determine the number of clusters. Among many cluster quality scores, we implement an adaptation

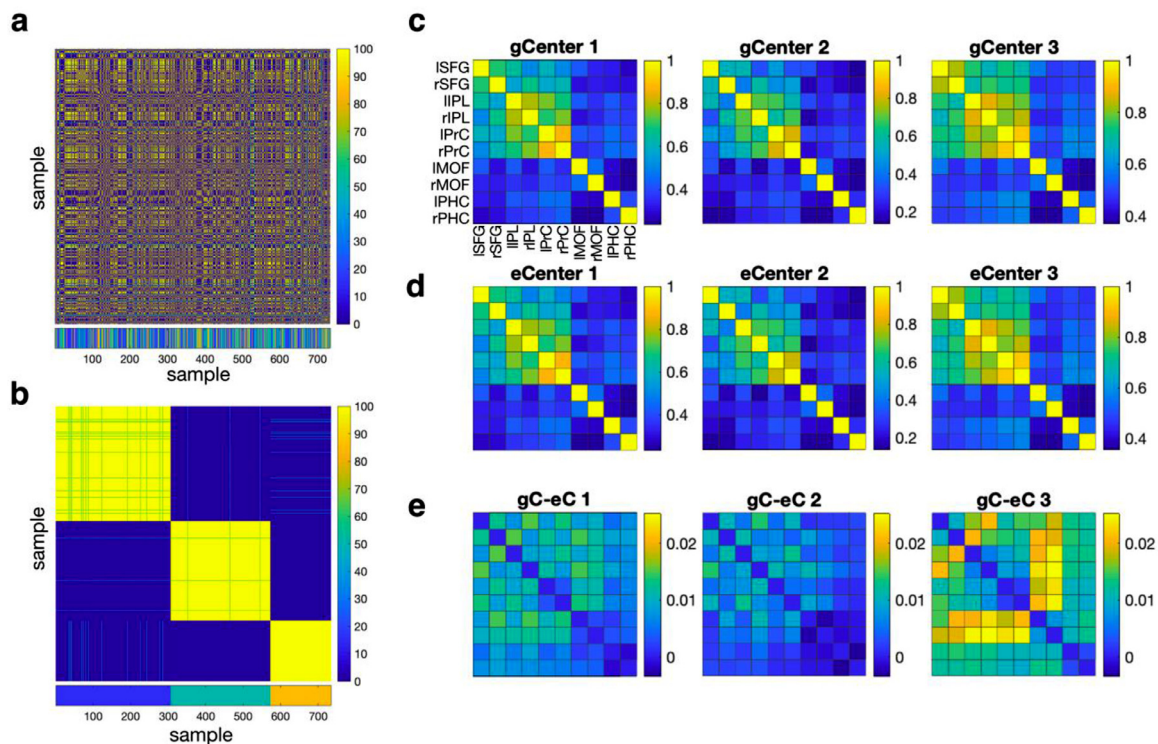


Fig. 16. Modularity optimization of multiple k -means clustering. Total 100 runs of k -means clustering compose a frequency adjacency matrix of all pairs of two nodes being in a same cluster (a). The modularity optimization of the frequency adjacency matrix results in sorted clusters (b). The centers for all three clusters derived by geodesic mean (c), Euclidean mean (d) and their difference (e) are presented.

of celebrated Silhouette method (Rousseeuw, 1987) to our context since it only relies on any distance metric.

To overcome the problems inherent in the k -means clustering such as dependency on the initial value and cluster number, we grafted a modularity-based clustering method (Jung et al., 2019) onto SPD space. This clustering approach constructs a frequency adjacency matrix that counts partition co-occurrence for each pair of samples after multiple runs of k -means clustering. The modularity optimization of the frequency adjacency matrix leads to reliable clustering of correlation matrices, robust to selection of initial values and cluster numbers.

To note, Cluster Validity Index is a thematic program to measure effectiveness of a given clustering though it has not been fully extended for Riemannian manifold. The extension of the popular evaluation indices from literature to large-scale clustering of functional connectivity matrices remains to be researched further.

Any SPD clustering algorithms can readily be used to the analysis of dynamic functional networks as shown in Fig. 17. The dynamics of functional connectivity has been summarized by transition matrix among several representative states, which was estimated by k -means clustering in the Euclidean space (Allen et al., 2014; Damaraju et al., 2014). We showed that this framework for dynamic connectivity analysis can be applied in the SPD space.

As mentioned earlier, a collection of correlation matrices takes account into a submanifold of SPD matrices with a unit diagonal constraint. This causes some operations with correlation matrices not necessarily resulting in another correlation matrix. In this study, we normalized resulting matrices into correlation matrices whenever necessary under the philosophy of iterative projection onto the submanifold, which is a constrained set within SPD manifold. If we consider one convention to normalize BOLD signals to rule out the effect of signals' magnitude, this inevitably converts covariance to correlation matrices while the literature has been little careful in distinguishing unexpected impact of such scenario. At this stage, we resort to an ad-hoc approach to deal with correlation matrices even within the SPD geometric structure. One

candidate to overcome such inconvenience includes ellipsope structure even though current literature largely deals with rank-deficient correlation matrix and lacks concrete theoretical studies at the moment. For the full-rank scenario, it is possible to consider a Riemannian submersion by an isometric action by the set of positive diagonal matrices from the Lie group theory to extend the affine-invariant framework. However, this exponentiates the computational burden while the gain is thin compared to that of our ad-hoc approach. This necessitates to further extensive investigation of what we would benefit from adopting such structure.

Statistical analysis or classification of functional connectivity over the SPD manifold is recommended compared to directly applying those methods to vectorized correlation matrices because the inter-relatedness among elements in the covariance matrices violates “the uncorrelated feature assumption” implicit in most statistical analysis or classification algorithms (Ng et al., 2016). In practice, however, analysis of functional connectivity under SPD framework may not always be beneficial compared to the Euclidean treatment in statistical analysis or in the machine learning. This may be particularly true in the high dimensional networks which incurs high computational cost. One possible bottleneck comes from an observation that in both AIRM and LERM geometries, an atomic operation that is no more algorithmically divisible involves matrix exponential and logarithm which are of comparable complexity as eigen-decomposition of symmetric matrices. Still, it does not necessarily hamper application of geometric approaches as long as one avoids analysis at the scale larger than tens of thousands of variables with modern computing capabilities. Although this is not usual in the general neuroimaging researches (mostly less than 500 ROIs – it takes about 300 s to calculate geometric mean of 1000 covariances matrices with a size of 500 as shown in Fig. 18), executing aforementioned operations requires heavy exploitation of computing hardware, efficient design of memory use, and specialized libraries designed for scalable computation, which is an interesting research direction of its own importance in the future.

We would also like to point out one requirement for application of the introduced framework. When the number of scans is small or not

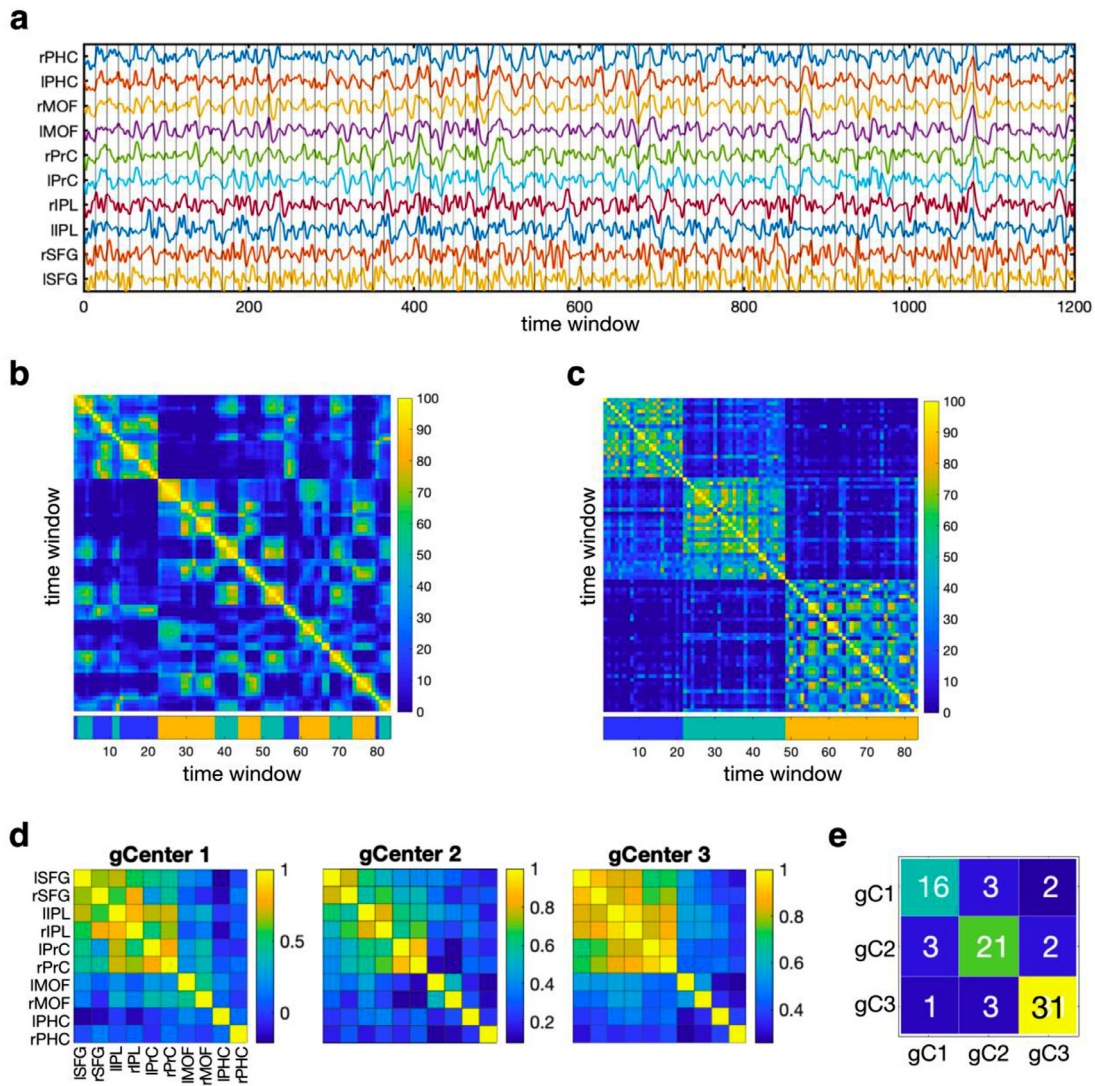


Fig. 17. State transition analysis of dynamic functional connectivity. A rs-fMRI time series of 1200 samples was divided into 83 consecutive epochs (a window size of 42 samples with 28 samples of overlap) (a) and correlation matrices were calculated from each epoch. The vertical lines indicate the starting point of each epoch (the gap is 28 samples). The 83 correlation matrices were clustered into three clusters using the modularization optimization of frequency adjacency matrix (b) from 100 runs of *k*-means clustering. The sorted frequency adjacency matrix is presented in (c). The centers for the three modules were evaluated by geodesic average of the connectivity matrices of the three modules (d). The state transition matrix among the three states (connectivity matrix) is presented in (e).

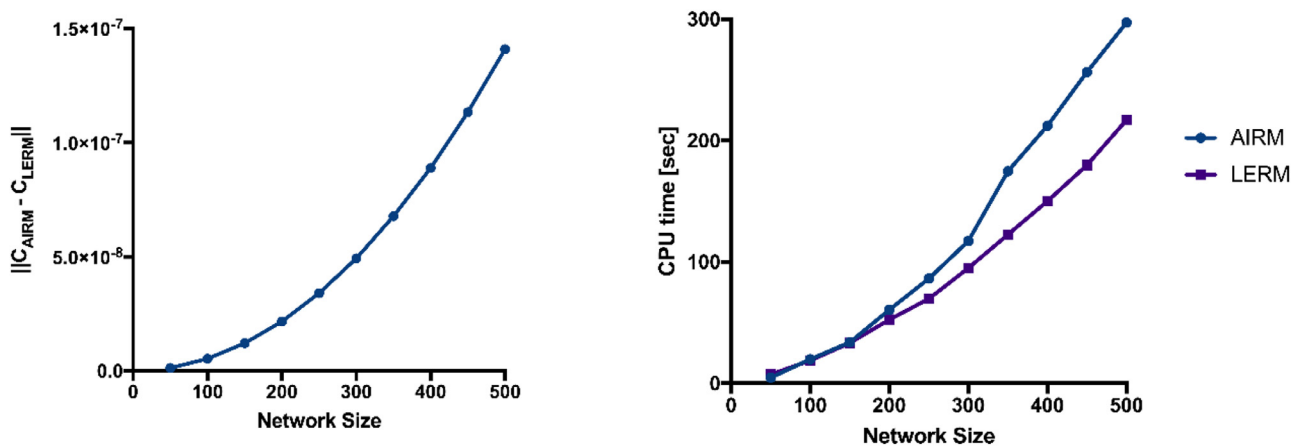


Fig. 18. Comparison between AIRM and LERM. The mean squared difference between Fréchet mean of 1000 generated covariances estimated by using AIRM and LERM with different matrix sizes (left) and the average computational time (in seconds) to compute Fréchet mean of 1000 covariance matrices with different matrix sizes (right).

sufficient enough compared to that of variables, empirical covariance, precision, and correlation matrices (conventionally used for fMRI connectivity analysis) are likely to be rank-deficient while two geometries we revisited require those representation matrices to be strictly positive definite. To mitigate this issue, some of the aforementioned frameworks such as Wasserstein, Procrustes, Cholesky, and any other geometries that do not require operations like matrix exponential and logarithm can be an alternative (Dryden et al., 2009; Takatsu, 2011). Also, it is natural to consider using more elaborated methods to estimate such matrices under assumptions like sparsity, strict non-singularity, and others. Estimation of high-dimensional covariance and precision matrices has long been a major problem in statistics and related communities in that we refer interested readers to Fan et al. (2016), Hong and Kim (2018), Lam (2019) for reviews to explore deep and expanded line of research on the topic.

The geometric approach to functional connectivity analysis is still early in the neuroimaging fields and we expect to find more examples that show advantages of dealing correlation matrices on the SPD space. We also expect that a MATLAB based toolbox for analyzing connectivity matrices would expediate geometric analysis of functional networks on the SPD manifold.

Acknowledgement

This research was supported by Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C7A1049051) and (NRF-2017M3C7A1030750). Authors thank Dr. Seok-Oh Jeong and Dr. Lizhen Lin for discussion.

Authors contributions

Kisung You: Conceptualization, Methodology, Software, Writing and Editing

Hae-Jeong Park: Conceptualization, Methodology, Validation, Data Curation, Writing and Editing, Funding

References

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Boardman, J.P., Murgasova, M., Doria, V., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2010. Combining morphological information in a manifold learning framework: application to neonatal MRI. *Med. Image Comput. Assist. Int.* 13, 1–8.

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D., 2014. Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676.

Aloise, D., Deshpande, D., Hansen, P., Papat, P., 2009. NP-hardness of Euclidean sum-of-squares clustering. *Machine Learn.* 75, 245–248.

Arsigny, V., Fillard, P., Pennec, X., Ayache, N., 2007. Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices. *SIAM J. Matrix Anal. Appl.* 29, 328–347.

Arthur, D., Vassilvitskii, S., 2007. *K-means++: The Advantages of Careful Seeding*. Society for Industrial and Applied Mathematics. Philadelphia, PA, USA, pp. 1027–1035.

Bhatia, R., 2015. *Positive Definite Matrices*. Princeton University Press.

Bhattacharya, A., Bhattacharya, R.N., 2012. *Nonparametric Inference on Manifolds: with Applications to Shape Spaces*. Cambridge University Press, Cambridge, UK New York.

Bilker, W.B., Hansen, J.A., Brensing, C.M., Richard, J., Gur, R.E., Gur, R.C., 2012. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* 19, 354–369.

Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541.

Biswas, M., Ghosh, A.K., 2014. A nonparametric two-sample test applicable to high dimensional data. *J. Multivariate Anal.* 123, 160–171.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. *Ann. N Y Acad. Sci.* 1124, 1–38.

Buyse, D.J., Reynolds 3rd, C.F., Monk, T.H., Berman, S.R., Kupfer, D.J., 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 28, 193–213.

Calhoun, V.D., Miller, R., Pearlson, G., Adali, T., 2014. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84, 262–274.

Celebi, M.E., Kingravi, H.A., Vela, P.A., 2013. A comparative study of efficient initialization methods for the *k*-means clustering algorithm. *Expert Syst. Appl.* 40, 200–210.

Chang, C., Glover, G.H., 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50, 81–98.

Cherian, A., Sra, S., Banerjee, A., Papanikolopoulos, N., 2013. Jensen-Bregman LogDet Divergence with Application to Efficient Similarity Search for Covariance Matrices. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2161–2174.

Cho, H., Kim, C.H., Knight, E.Q., Oh, H.W., Park, B., Kim, D.G., Park, H.J., 2017. Changes in brain metabolic connectivity underlie autistic-like social deficits in a rat model of autism spectrum disorder. *Sci. Rep.* 7, 13213.

Cribben, I., Haraldsdottir, R., Atlas, L.Y., Wager, T.D., Lindquist, M.A., 2012. Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage* 61, 907–920.

Dai, M., Zhang, Z., Srivastava, A., 2016. Testing Stationarity of Brain Functional Connectivity Using Change-Point Detection in fMRI Data. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Las Vegas, NV, USA.

Dai, M., Zhang, Z., Srivastava, A., 2020. Analyzing Dynamical brain functional connectivity as trajectories on space of covariance matrices. *IEEE Trans. Med. Imaging* 39, 611–620.

Damaraju, E., Allen, E.A., Belger, A., Ford, J.M., McEwen, S., Mathalon, D.H., Mueller, B.A., Pearlson, G.D., Potkin, S.G., Preda, A., Turner, J.A., Vaidya, J.G., van Erp, T.G., Calhoun, V.D., 2014. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *Neuroimage Clin* 5, 298–308.

Davoudi, A., Ghidary, S.S., Sadatnejad, K., 2017. Dimensionality reduction based on distance preservation to local mean for symmetric positive definite matrices and its application in brain-computer interfaces. *J. Neural Eng.* 14, 036019.

Deligianni, F., Varoquaux, G., Thirion, B., Robinson, E., Sharp, D.J., Edwards, A.D., Rueckert, D., 2011. A probabilistic framework to infer brain functional connectivity from anatomical connections. *Inf. Process. Med. Imaging* 22, 296–307.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.

Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett Jr., J.R., Barch, D.M., Petersen, S.E., Schlaggar, B.L., 2010. Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361.

Dryden, I.L., Koloydenko, A., Zhou, D., 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* 3, 1102–1123.

Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R.N., Zebley, B., Oathes, D.J., Etkin, A., Schatzberg, A.F., Sudheimer, K., Keller, J., Mayberg, H.S., Gunning, F.M., Alexopoulos, G.S., Fox, M.D., Pascual-Leone, A., Voss, H.U., Casey, B.J., Dubin, M.J., Liston, C., 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* 23, 28–38.

Fan, J., Liao, Y., Liu, H., 2016. An overview of the estimation of large covariance and precision matrices. *Econo. J.* 19, C1–C32.

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671.

Fletcher, P.T., Joshi, S., 2004. Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors. In: *Proceedings of the MMBIA 2004, CVAMIA 2004*. Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, Lecture Notes in Computer Science. Springer, pp. 87–98.

Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S., 2004. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging* 23, 995–1005.

Ginestet, C.E., Li, J., Balachandran, P., Rosenberg, S., Kolaczky, E.D., 2017. Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* 11, 725–750.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., Consortium, W.U.-M.H., 2013. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124.

Goh, A., Vidal, R., 2008. Clustering and dimensionality reduction on Riemannian manifolds. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Anchorage, AK, USA.

Grubišić, I., Pietersz, R., 2007. Efficient rank reduction of correlation matrices. *Linear Algebra Appl.* 422, 629–653.

Handwerker, D.A., Roopchansingh, V., Gonzalez-Castillo, J., Bandettini, P.A., 2012. Periodic changes in fMRI connectivity. *Neuroimage* 63, 1712–1719.

Hong, Y., Kim, C., 2018. Recent developments in high dimensional covariance estimation and its related issues, a review. *J. Korean Statist. Soc.* 47, 239–247.

Hotelling, H., 1936. Relations Between Two Sets of Variates. *Biometrika* 28, 321.

Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattner, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tyagar, D., Vardi, M.Y., Weikum, G., Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., Thirion, B., 2010. Detection of Brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 200–208.

Hutchison, R.M., Womelsdorf, T., Gati, J.S., Everling, S., Menon, R.S., 2013. Resting-state networks show dynamic functional connectivity in Awake humans and anesthetized macaques. *Hum. Brain Mapp* 34, 2154–2177.

Jang, C., Knight, E.Q., Pae, C., Park, B., Yoon, S.A., Park, H.J., 2017. Individuality manifests in the dynamic reconfiguration of large-scale brain networks during movie viewing. *Sci. Rep.* 7, 41414.

Jang, C., Park, H.J., Chang, W.S., Pae, C., Chang, J.W., 2016. Immediate and longitudinal alterations of functional networks after thalamotomy in essential tremor. *Front. Neurosci.* 10, 184.

- Jeong, S.O., Pae, C., Park, H.J., 2016. Connectivity-based change point detection for large-size functional networks. *Neuroimage* 143, 353–363.
- Jung, K., Kang, J., Chung, S., Park, H.J., 2019. Dynamic causal modeling for calcium imaging: Exploration of differential effective connectivity for sensory processing in a barrel cortical column. *Neuroimage*, 116008.
- Kim, J., Wozniak, J.R., Mueller, B.A., Shen, X., Pan, W., 2014. Comparison of statistical tests for group differences in brain functional networks. *Neuroimage* 101, 681–694.
- Kyeong, S., Kim, E., Park, H.J., Hwang, D.U., 2014. Functional network organizations of two contrasting temperament groups in dimensions of novelty seeking and harm avoidance. *Brain Res.* 1575, 33–44.
- Lam, C., 2019. High-Dimensional Covariance Matrix Estimation. Wiley Interdisciplinary Reviews: Computational Statistics.
- Lee, D., Pae, C., Lee, J.D., Park, E.S., Cho, S.R., Um, M.H., Lee, S.K., Oh, M.K., Park, H.J., 2017. Analysis of structure-function network decoupling in the brain systems of spastic diplegic cerebral palsy. *Hum. Brain Mapp* 38, 5292–5306.
- Lehmann, E.L., Romano, J.P., 2010. Testing Statistical Hypotheses, 3 ed. Springer, New York, NY.
- Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.M., Schlupe, M., Vuilleumier, P., Van De Ville, D., 2013. Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *Neuroimage* 83, 937–950.
- Lin, L., St. Thomas, B., Zhu, H., Dunson, D.B., 2017. Extrinsic local regression on manifold-valued data. *J. Am. Statist. Assoc.* 112, 1261–1273.
- Lloyd, S., 1982. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* 28, 129–137.
- Maa, J.-F., Pearl, D.K., Bartoszyński, R., 1996. Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *Annals Stat.* 24, 1069–1074.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Moakher, M. (Ed.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1. Statistics. University of California Press, Berkeley, Calif., pp. 281–297.
- Moakher, M., 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* 26, 735–747.
- Monti, R.P., Hellyer, P., Sharp, D., Leech, R., Anagnostopoulos, C., Montana, G., 2014. Estimating time-varying brain connectivity networks from functional MRI time series. *Neuroimage* 103, 427–443.
- Nadaraya, E.A., 1964. On estimating regression. *Theory Probab. Appl.* 9, 141–142.
- Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. *Phys. Rev. E* 69.
- Ng, B., Dressler, M., Varoquaux, G., Poline, J.B., Greicius, M., Thirion, B., 2014. Transport on Riemannian manifold for functional connectivity-based classification. *Med. Image Comput. Assist. Interv.* 17, 405–412.
- Ng, B., Varoquaux, G., Poline, J.B., Greicius, M., Thirion, B., 2016. Transport on Riemannian Manifold for Connectivity-Based Brain Decoding. *IEEE Trans. Med. Imaging* 35, 208–216.
- Park, B., Kim, D.S., Park, H.J., 2014. Graph independent component analysis reveals repertoires of intrinsic network components in the human brain. *PLoS One* 9, e82873.
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space the London, Edinburgh, and Dublin philosophical magazine. *J. Sci.* 2, 559–572.
- Pennec, X., 2006. Intrinsic statistics on riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vis.* 25, 127.
- Pennec, X., Fillard, P., Ayache, N., 2006. A riemannian framework for tensor computing. *Int. J. Comput. Vision* 66, 41–66.
- Pitman, E.J.G., 1937. Significance tests which may be applied to samples from any populations supplement. *J. R. Stat. Soc.* 4, 119.
- Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional network organization of the human brain. *Neuron* 72, 665–678.
- Preti, M.G., Bolton, T.A., Van De Ville, D., 2017. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage* 160, 41–54.
- Qiu, A., Lee, A., Tan, M., Chung, M.K., 2015. Manifold learning on brain functional networks in aging. *Med. Image Anal.* 20, 52–60.
- Rahim, M., Thirion, B., Varoquaux, G., 2019. Population shrinkage of covariance (PoSCE) for better individual brain functional-connectivity estimation. *Med. Image Anal.* 54, 138–148.
- Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069.
- Sadatnejad, K., Rahmati, M., Rostami, R., Kazemi, R., Ghidary, S.S., Muller, A., Alimardani, F., 2019. EEG representation using multi-instance framework on the manifold of symmetric positive definite matrices. *J. Neural Eng.* 16, 036016.
- Siman-Tov, T., Bosak, N., Sprecher, E., Paz, R., Eran, A., Aharon-Peretz, J., Kahn, I., 2016. Early age-related functional connectivity decline in high-order cognitive networks. *Front Aging Neurosci.* 8, 330.
- Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E., Glasser, M.F., Ugurbil, K., Barch, D.M., Van Essen, D.C., Miller, K.L., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* 18, 1565–1567.
- Takatsu, A., 2011. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* 48, 1005–1026.
- Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., Consortium, W.U.-M.H., 2012. The Human Connectome Project: a data acquisition perspective. *Neuroimage* 62, 2222–2231.
- Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., Thirion, B., 2010. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. *Med. Image Comput. Assist. Interv.* 13, 200–208.
- Wang, Z., Vemuri, B.C., 2005. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Trans. Med. Imaging* 24, 1267–1277.
- Wasserman, L., 2006. *All of Nonparametric Statistics*. Springer, New York.
- Watson, G.S., 1964. Smooth regression analysis. *Sankhyā: Ind. J. Stat. Series A* (1961-2002) 26, 359–372.
- Xie, X., Yu, Z.L., Lu, H., Gu, Z., Li, Y., 2017. Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 504–516.
- Yahata, N., Morimoto, J., Hashimoto, R., Lisi, G., Shibata, K., Kawakubo, Y., Kuwabara, H., Kuroda, M., Yamada, T., Megumi, F., Imamizu, H., Nanez Sr., J.E., Takahashi, H., Okamoto, Y., Kasai, K., Kato, N., Sasaki, Y., Watanabe, T., Kawato, M., 2016. A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat Commun.* 7, 11254.
- Yamin, A., Dayan, M., Squarcina, L., Brambilla, P., Murino, V., Diwadkar, V., Sona, D., 2019. Comparison of brain connectomes using geodesic distance on manifold: a twin study. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, pp. 1797–1800.
- Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165.