



OPEN

Geometric learning of functional brain network on the correlation manifold

Kisung You^{1,2} & Hae-Jeong Park^{2,3,4}✉

The correlation matrix is a typical representation of node interactions in functional brain network analysis. The analysis of the correlation matrix to characterize brain networks observed in several neuroimaging modalities has been conducted predominantly in the Euclidean space by assuming that pairwise interactions are mutually independent. One way to take account of all interactions in the network as a whole is to analyze the correlation matrix under some geometric structure. Recent studies have focused on the space of correlation matrices as a strict subset of symmetric positive definite (SPD) matrices, which form a unique mathematical structure known as the Riemannian manifold. However, mathematical operations of the correlation matrix under the SPD geometry may not necessarily be coherent (i.e., the structure of the correlation matrix may not be preserved), necessitating a post-hoc normalization. The contribution of the current paper is twofold: (1) to devise a set of inferential methods on the correlation manifold and (2) to demonstrate its applicability in functional network analysis. We present several algorithms on the correlation manifold, including measures of central tendency, cluster analysis, hypothesis testing, and low-dimensional embedding. Simulation and real data analysis support the application of the proposed framework for brain network analysis.

One of the widely accepted perspectives on the human brain claims that the brain is a network consisting of interactions among distributed regions¹. Such interactions are often captured by the correlation matrix of spontaneous fluctuations, observed in the resting-state functional magnetic resonance imaging (rs-fMRI)² or electroencephalogram (EEG) / magnetoencephalogram (MEG)^{3,4}. Representations of functional brain network as a correlation matrix have been central parts of analysis in diverse fields of brain research such as exploration of brain diseases⁵⁻⁷, individual identification⁸, between-individual variability⁹, and dynamic connectivity analysis¹⁰⁻¹⁴ to name a few. Studying with representation matrices are typically done by considering interactions (i.e., edges of the network) independently^{15,16} or together to explain certain dependent edges^{17,18}.

One limitation of the aforementioned approaches is that it does not take an intrinsic dependence structure among edges into consideration. Mathematically, the correlation matrix is a subtype of symmetric and positive-definite (SPD) matrices, whose collection forms a unique geometric structure called Riemannian manifolds. In neuroimaging communities, SPD approach has gained popularity with successful applications such as exploratory analysis of group-level functional connectivity^{19,20}, statistical hypothesis testing²¹, regression for functional-structural prediction²², individual fingerprinting²³, and so on.

The geometric approach of SPD matrices has also been applied to studies with correlation matrices¹⁹. However, the naive SPD-based approach incurs a practical issue that cannot be easily neglected. Suppose we are given two correlation matrices C_1 and C_2 . Under a variety of geometric characterizations of SPD space, a mean of two matrices may not be a correlation matrix, either numerically or analytically, except for the Euclidean geometry. Since the space of correlation matrices is a strict subset of the SPD manifold, the unsophisticated application of SPD geometry to correlation matrices does not guarantee any valid operations on the manifold to result in a correlation matrix. Therefore, it is natural to consider a more strict geometric structure regarding studies with a set of correlation matrices.

Compared to the SPD manifold, the space of correlation matrices known as elliptope²⁴ has little been studied except for a few notable works that lack desirable properties or availability of computational routines^{25,26}. In

¹Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA. ²Center for Systems and Translational Brain Science, Institute of Human Complexity and Systems Science, Yonsei University, Seoul, Republic of Korea. ³Graduate School of Medical Science, Brain Korea 21 Project, Department of Nuclear Medicine, Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea. ⁴Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea. ✉email: parkhj@yonsei.ac.kr

response, a recent study²⁷ proposed to use a quotient geometry of SPD manifold induced by affine-invariant Riemannian metric (AIRM)²⁸ for the space of correlation matrices. In a layperson's language, the induced manifold structure of correlation space inherits characteristics of the ambient SPD manifold to allow similar machinery for inference on the correlation manifold. Also, core numerical routines on the correlation manifold were elaborated under the quotient geometry²⁹.

In this work, we incorporate the recent development of the quotient geometry of the correlation manifold into well-known algorithms in machine learning and statistical inference for functional connectivity analysis. The proposed algorithms over the correlation manifold include central tendency measures, cluster analysis, hypothesis testing, and low-dimensional embedding. We then demonstrate the efficacy of the proposed framework using simulated and real data examples. All the algorithms are implemented as a MATLAB toolbox (Mathworks, inc. USA) for public use, not only for functional network analysis but also for any application fields using the correlation matrix on the proper geometry.

Methods

Quotient geometry of the correlation manifold. Our main interest is to study the space of correlation matrices, which are symmetric, positive-definite, and have unit diagonal elements. We denote a space of $(n \times n)$ correlation matrices as $\mathcal{C}_{++}^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^T, \text{rank}(X) = n, \text{diag}(X) = 1\}$, which is of dimension $n(n-1)/2$. From the definition, we see that the space of correlation matrices is a strict subset of the SPD manifold denoted as \mathcal{S}_{++}^n . David²⁷ employed a well-known result from Lie group theory³⁰ to equip \mathcal{C}_{++}^n with manifold structure. Theorem 2.1 of²⁷ showed that the Lie group \mathcal{D}_{++}^n of strictly positive entries acts smoothly, properly, and freely on \mathcal{S}_{++}^n so that the quotient manifold $\mathcal{S}_{++}^n/\mathcal{D}_{++}^n$ is a smooth manifold that is uniquely expressed by correlation matrices. Therefore, any SPD matrix $\Sigma \in \mathcal{S}_{++}^n$ can be mapped to the correlation manifold by an invariant submersion $\text{diag}(\Sigma)^{-1/2} \cdot \Sigma \cdot \text{diag}(\Sigma)^{-1/2} \in \mathcal{C}_{++}^n \cong \mathcal{S}_{++}^n/\mathcal{D}_{++}^n$. Furthermore, a metric on \mathcal{C}_{++}^n is induced by any metric on the ambient space \mathcal{S}_{++}^n when invariant under the group action of \mathcal{D}_{++}^n . This enables the geometry of \mathcal{C}_{++}^n to inherit characteristics of an ambient SPD manifold under the affine-invariant Riemannian metric in many aspects. We refer interested readers to Supplementary Information for more details on elements of Riemannian geometry and construction of geometric structure along with core computational routines on the correlation manifold.

Measures of central tendency. We now introduce several algorithms for learning and inference a given set of correlation matrices. For the rest of this paper, we use the following notations; $\{C_i\}_{i=1}^N \subset \mathcal{C}_{++}^n$ for N observations of $(n \times n)$ correlation matrices, μ and μ_j for means of all data and class j , S_i an index set of observations that belong to the cluster i , superscript in parenthesis (t) for indexing an iteration t , and $d: \mathcal{C}_{++}^n \times \mathcal{C}_{++}^n \rightarrow \mathbb{R}_+$ a distance function.

One of the primary characterizations for an empirical distribution of the data is measures of central tendency such as mean, median, and mode. For manifold-valued data, however, such entities are not trivially obtained due to the nonlinear nature of a Riemannian manifold. The Fréchet mean, also known as Riemannian center of mass, is a generalization of the aforementioned concepts to arbitrary metric spaces³¹. More generally, the sample version of the Riemannian L_p center of mass μ_p is defined as a minimizer of the functional $f(\mu)$,

$$\mu_p = \underset{\mu \in \mathcal{C}_{++}^n}{\text{argmin}} f(\mu) \text{ where } f(\mu) = \frac{1}{N} \sum_{i=1}^N d^p(\mu, C_i),$$

for $1 \leq p < \infty$ ³². Given a minimizer $\hat{\mu}_p$, the sample variation $V_p = \sum_{i=1}^N d^p(\hat{\mu}_p, C_i)/N$ quantifies dispersion of the distribution. For example, when $p = 2$ and data lies on \mathbb{R} , this corresponds to the sample variance.

For the special cases of $p = 1$ and $p = 2$, the minimizers are also known as Fréchet median and Fréchet mean, respectively³³. For the Fréchet mean computation, one of the standard algorithms is Riemannian gradient descent²⁸ which we adopt in our implementation (**corr_mean.m**). At the t -step, gradient of the cost function $f(\mu)$ is evaluated on the tangent space of $\mu^{(t)}$ and projected back onto \mathcal{C}_{++}^n via exponential map. To summarize, with an initial point $\mu^{(0)}$, we repeat the following steps to obtain the sample Fréchet mean

$$\begin{aligned} \nabla f^{(t)} &= \frac{2}{N} \sum_{i=1}^N \log_{\mu^{(t)}}(C_i), \\ \mu^{(t+1)} &= \exp_{\mu^{(t)}}(-\nabla f^{(t)}), \end{aligned}$$

until convergence. A convenient choice of stopping criterion is to iterate until $\|\mu^{(t+1)} - \mu^{(t)}\|_F < \varepsilon$ for a specified tolerance level ε . This criterion stops the algorithm by a small increment in Frobenius norm rather than the geodesic distance. We recommend this criterion for two reasons. First, computing the incremental change in Frobenius norm is much cheaper than the geodesic distance. Second, a small increment implies that the evaluated gradient is of very small magnitude so that an iterate is sufficiently close to a critical point of the functional.

The function **corr_median.m** computes sample Fréchet median using a Riemannian adaptation of the Weiszfeld algorithm^{34–37}. The cost function for minimization is as follows

$$g(\mu) = \frac{1}{N} \sum_{i=1}^d (\mu, C_i) \text{ for some } \mu \in \mathcal{C}_{++}^n.$$

Given an initial point $\mu^{(0)}$, the Riemannian Weiszfeld algorithm iterates the following steps

$$W^{(t)} = \frac{\sum_{i=1}^N w_i^{(t)} \log_{\mu^{(t)}}(C_i)}{\sum_{i=1}^N w_i^{(t)}},$$

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

for weights $w_i^{(t)} = 1/d(\mu^{(t)}, C_i)$, $i = 1, \dots, N$. It is worthy to mention a case where an iterate $\mu^{(t)}$ coincides with one of the observations C_i 's, which incurs the singularity of the corresponding weight. Common strategies to avoid such issues may include stopping an algorithm at convergence to one of the data points, adjusting a weight by adding a sufficiently small number, or removing the coinciding observations' contributions.

Cluster analysis. Next, we implement three clustering algorithms that partition the data into K disjoint subsets called clusters and two measures of cluster validity that help determine the optimal number of clusters K in a data-driven manner.

The k -means algorithm³⁸ is one of the primary methods for cluster analysis, which can be easily extended to Riemannian manifolds where a routine for computing the mean is readily available. We implemented a standard version of Lloyd's algorithm³⁹ that makes an iterative refinement of the cluster assignment (**corr_kmeans.m**).

1. Select K data points as cluster centroids $\{\mu_i^{(0)}\}_{i=1}^K$.
2. Repeat the following steps until convergence:
 - Compute distances from an observation to cluster centroids and make an assignment to the cluster with minimal distance. The cluster membership $S^{(t)} = [S_1^{(t)}, \dots, S_K^{(t)}]$ is given by

$$S_k^{(t)} = \{i \mid d(C_i, \mu_k^{(t)}) \leq d(C_i, \mu_j^{(t)}) \text{ for all } j \neq k\}.$$

When an observation belongs to multiple clusters, assign it randomly.

- For each cluster k , update cluster centroid by Fréchet mean of corresponding observations,

$$\mu_k^{(t+1)} = \operatorname{argmin}_{\mu \in \mathcal{L}_{++}^n} \sum_{i \in S_k^{(t)}} d^2(\mu, C_i).$$

The k -medoids algorithm⁴⁰ is another popular partitioning method. Compared to the k -means algorithm, k -medoids use a central observation of a given cluster as a centroid that minimizes the average dissimilarities. This characteristic enables the algorithm to be used on an arbitrary metric space, not to mention Riemannian manifolds. Replacing the mean with one of the observations gives two benefits. First, the k -medoids algorithm is regarded to be more robust to outliers⁴¹. More importantly, the algorithm does not require explicit computation of cluster means as it selects observations that play the role of central objects. This is especially appealing in our context since the computation of Fréchet means on the correlation manifold involves nested iterations. Our implementation (**corr_kmedoids.m**) employs that of the original partitioning around medoids algorithm (PAM)⁴⁰, which is similar to Lloyd's algorithm except that the update procedure for cluster centroids is performed by selecting the observations with minimal average dissimilarities.

The last clustering method we included is the spectral clustering algorithm^{42–44}. Inspired by spectral graph theory⁴⁵, spectral clustering first finds a low-dimensional embedding via the spectral information of a data-affinity graph and its Laplacian. Our implementation of the algorithm (**corr_specc.m**) makes use of data-adaptive construction for the affinity graph⁴⁶ and the normalized symmetric Laplacian matrix⁴⁴. The algorithm is described as follows.

1. Construct an $(N \times N)$ affinity/similarity matrix S ,

$$S_{ij} = \exp\left(-\frac{d(C_i, C_j)^2}{\sigma_i \cdot \sigma_j}\right),$$

where σ_i is the distance from C_i to its k -th nearest neighbor.

2. The graph Laplacian is defined as

$$L = D^{-1/2} S D^{-1/2},$$

where D is called a degree matrix with entries $D_{ii} = \sum_j S_{ij}$ and $D_{ij} = 0$ for $i \neq j$.

3. Denote $V \in \mathbb{R}^{N \times K}$ whose columns are K eigenvectors of L that correspond to K smallest eigenvalues.
4. Normalize each row of V by $V_i \leftarrow V_i / \|V_i\|$.
5. Cluster assignment is obtained by applying k -means clustering algorithm onto the rows of V .

When little prior knowledge or strong assumptions are available for the intrinsic clustering structure of the data, cluster validity indices offer a way to quantify how coherent the attained clustering assignment is regarding the data⁴⁷. Among many quality indices, we offer two of the celebrated indices, silhouette score (**corr_silhouette.m**) and Calinski-Harabasz (CH) index (**corr_CH.m**).

First, silhouette score⁴⁸ measures proximity of observations within a cluster to its neighboring clusters. For each observation C_i , a silhouette value is defined as $s(i) = \{b(i) - a(i)\}/\max\{a(i), b(i)\}$ for two auxiliary quantities

$$a(i) = \frac{1}{|S_i| - 1} \sum_{j \in S_i \setminus \{i\}} d(C_i, C_j) \quad \text{and} \quad b(i) = \min_{i \neq j} \frac{1}{|S_j|} \sum_{j \in S_j} d(C_i, C_j),$$

where S_i denotes the indices that share the same label as an i -th observation. The quantity $a(i)$ measures cohesiveness of a cluster by averaging distances between C_i and the rest within same cluster, while $b(i)$ quantifies minimal degree of separation from an observation to all points in other clusters. The global silhouette score $S^* = \sum_{i=1}^N s(i)/N$ is defined as an arithmetic mean of pointwise silhouette values lying in $[-1, 1]$. A partition is considered to be optimal for large values of the global silhouette score.

CH index⁴⁹ is represented as a ratio of degrees of separation and cohesion. For a partition of K disjoint clusters, denote $\{\mu_k\}_{k=1}^K$ and μ as Fréchet means per class and the entire dataset, respectively. The index is defined as follows,

$$CH^* = \frac{\text{degree of separation}}{\text{degree of cohesion}} = \frac{\sum_{k=1}^K |S_k| \cdot d(\mu_k, \mu)^2 / (K - 1)}{\sum_{k=1}^K \sum_{i \in S_k} d(C_i, \mu_k)^2 / (N - K)},$$

where the terms resemble those from discriminant analysis⁵⁰. As the definition of the index is a ratio, it is reasonable to consider a partition of higher CH index as a superior candidate for clustering as it indicates that the given partition describes the clustering structure of the data.

Dimension reduction for visualization. In order to visualize how the set of correlation matrices is distributed, we implemented three methods for dimension reduction; classical and metric multidimensional scaling (MDS)⁵¹ and principal geodesic analysis (PGA)⁵².

MDS refers to a class of algorithms for low-dimensional data embedding based on pairwise similarities. Among many variants, we present classical (**corr_cmds.m**) and metric (**corr_mmds.m**) variants of MDS. We denote $X_i \in \mathbb{R}^p, i = 1, \dots, N$ a set of low-dimensional embeddings in some Euclidean space and $D_{N \times N}$ a matrix of pairwise distances, i.e., $D_{ij} = d(C_i, C_j)$.

Classical MDS minimizes the cost function known as *strain*

$$\text{Strain}(X_{1:N}) = \sqrt{\frac{\sum_{ij} (B_{ij} - X_i^\top X_j)^2}{\sum_{ij} B_{ij}^2}},$$

for a doubly centered matrix B

$$B = -\frac{1}{2} \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right) D^{(2)} \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right),$$

where $D^{(2)}$ is an element-wise square of matrix D . We note that the minimizer of strain function is analytically available via eigen-decomposition of the doubly centered matrix B and the embedding thereof is identical to that of PCA.

One limitation of classical MDS is that the method assumes the data dissimilarity D be given by standard L_2 norm in Euclidean space, which diminishes the interpretability of the attained embedding. Metric MDS provides an alternative approach by minimizing the stress function

$$\text{Stress}(X_{1:N}) = \sqrt{\sum_{i \neq j} (D_{ij} - \|X_i - X_j\|)^2}.$$

The method aims at finding an optimal embedding that maximally preserves the pairwise distance structure regardless of the original metric space. As no closed-form solution exists for the stress minimization, we use the SMACOF algorithm⁵³ that optimizes the cost objective by finding a majorizing function and updating an iterate by a minimizer of the majorizing function until convergence.

PGA is an adaptation of the principal component analysis (PCA)⁵⁴ to the manifold setting. The core idea of PGA is to utilize the property of tangent space as a vector space. We briefly describe the procedure as follows. Once the Fréchet mean $\hat{C} \in \mathcal{C}_{++}^n$ is computed, every observation C_1, \dots, C_N is projected onto the tangent space $T_{\hat{C}} \mathcal{C}_{++}^n$ using the logarithm map $U_i = \log_{\hat{C}} C_i$. The native version of PGA constructs an empirical covariance matrix $\Sigma = \sum_{i=1}^N U_i U_i^\top / N$ onto which eigen-decomposition is applied. If we denote eigen-pairs as $\{\lambda_i, v_i\}$ for $i = 1, \dots, \dim(\mathcal{C}_{++}^n)$, local coordinates for an observation i are given by $y_i = [v_1 \dots v_k]^\top \log_{\hat{C}}(C_i) \in \mathbb{R}^k$. As the name suggests, PGA considers loadings to be projected basis onto the manifold via exponential map $\{\exp_{\hat{C}}(v_i)\}$. Unlike conventional PCA on Euclidean space, the loadings, also known as principal geodesics, are not orthogonal in that they should be regarded as an approximate basis in the vicinity of the Fréchet mean.

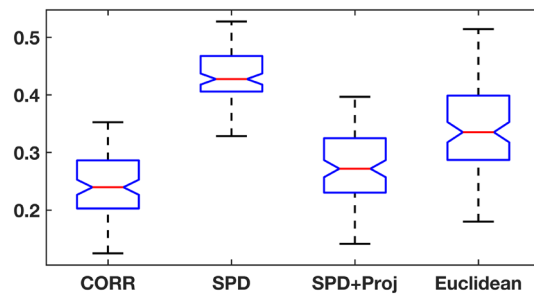


Figure 1. Comparison of error distributions for means of simulated correlation matrices under different geometries.

Results

Simulation 1: Fréchet mean. In order to see whether the quotient geometry of the correlation manifold is a convincing alternative, we compared the effectiveness of Fréchet means under different geometries. We assume a simple multivariate Gaussian model $\mathcal{N}(0_5, I_5)$ in \mathbb{R}^5 as a generating process of the data and draw 50 observations for sample correlation computation to assure positive definiteness. At each iteration, 20 sample correlation matrices C_1, \dots, C_{20} are generated and different means are obtained, including (1) Fréchet mean on \mathcal{C}_{++}^n , (2) Fréchet mean on \mathcal{S}_{++}^n with AIRM geometry and (3) its projection onto \mathcal{C}_{++}^n by a submersion Π , and (4) Euclidean mean $\sum_{i=1}^{20} C_i/20$. We recall that the third approach to combine SPD geometry and projection was extensively tested for functional connectivity analysis and showed good results on several inferential tasks⁵⁵.

Figure 1 shows empirical error distribution of different means against the identity matrix measured in the Frobenius norm. The quotient geometry for the correlation matrix shows the least degree of error while AIRM performed poorly. This phenomenon is not surprising since the Fréchet mean of correlation matrices under AIRM is, in general, not even a correlation matrix. On the other hand, the post-hoc projection of SPD mean estimate onto \mathcal{C}_{++}^n shows compatible results with that of the correlation manifold structure.

We then compared the Fréchet mean and the Euclidean average. We generated 30 samples from three model correlation matrices by adding noise with a standard deviation of 1 in the geometric tangent space. The model correlation matrices were derived from human connectivity matrices obtained from resting-state fMRI. These matrices are mutually distant from each other under all of Euclidean, SPD, and correlation geometries. Figure 2a shows five samples derived from the three model correlation matrices, which are presented in the first row of Fig. 2b. The second and fourth rows indicate the Fréchet mean and the Euclidean average (mean of each element of the correlation matrix). The third and fifth rows indicate the difference between the model correlation matrix and the Fréchet mean and the Euclidean average.

Next, we compared wall-clock time of Fréchet mean computation under correlation and SPD geometry of AIRM with varying number of observations and dimensionality. The results are summarized in Table 1, where correlation structure puts a more immense computational burden than AIRM. This result is due to the logarithmic map on the correlation manifold, where a minimization problem over \mathcal{D}_{++}^n needs to be solved for every observation at each iteration. One remedy is to limit the number of iterations for an internal optimization problem at the cost of using a suboptimal diagonal matrix. We empirically witnessed that the approximate solution did not sacrifice the performance much while saving a considerable amount of time.

Simulation 2: dimension reduction and cluster analysis. In this example, we generated 90 correlation matrices consisting of three populations of size 30 for each of the three model matrices used above (C1, C2, and C3) in \mathcal{C}_{++}^5 .

First, we applied three dimension reduction algorithms, i.e., classical multidimensional scaling (CMDS), metric multidimensional scaling (MMDS), and principal geodesic analysis (PGA) to the generated data from the three model correlation matrices. Embeddings shown in Fig. 3 assert validity of the presented algorithms.

Next, we validate three clustering algorithms on the generated data with varying numbers of clusters K . Since the true number of clusters in this simulation is 3, a reliable clustering algorithm is expected to return a coherent partition when $K = 3$ and discourage others. We summarized the results in Table 2 where the Silhouette score and CH index are used to quantify the effectiveness of clustering algorithms under the choice of K . It is not surprising that all methods showed the pattern that a peak appears for $K = 3$ while discriminating $K = 2$ and $K = 4$ since the data is well separated, as shown previously.

Simulation 3: hypothesis testing. We tested the effectiveness of two nonparametric tests for equality of distributions. For model correlation matrices C1, C2, and C3, we generated 100 perturbed observations and applied the two tests in a pairwise manner on three samples. For each setting, we ran 10^5 permutations to make the Monte Carlo procedure credible. In all pairwise comparisons, both tests showed extremely significant p values of 10^{-5} . This means that every resampled statistic was smaller than that of the data $\hat{T}_{m,n}$. This implies that both tests can distinguish two well-separated empirical measures on the correlation manifold.

Next, we visualized an empirical distribution of p values under the null hypothesis of equal distributions, which is known to follow a uniform distribution on $[0, 1]$ ⁵⁶. For a model correlation matrix C_1 , two sample sets were drawn with noise, each consisting of 30 observations. Figure 4 shows the histogram of p values from 200

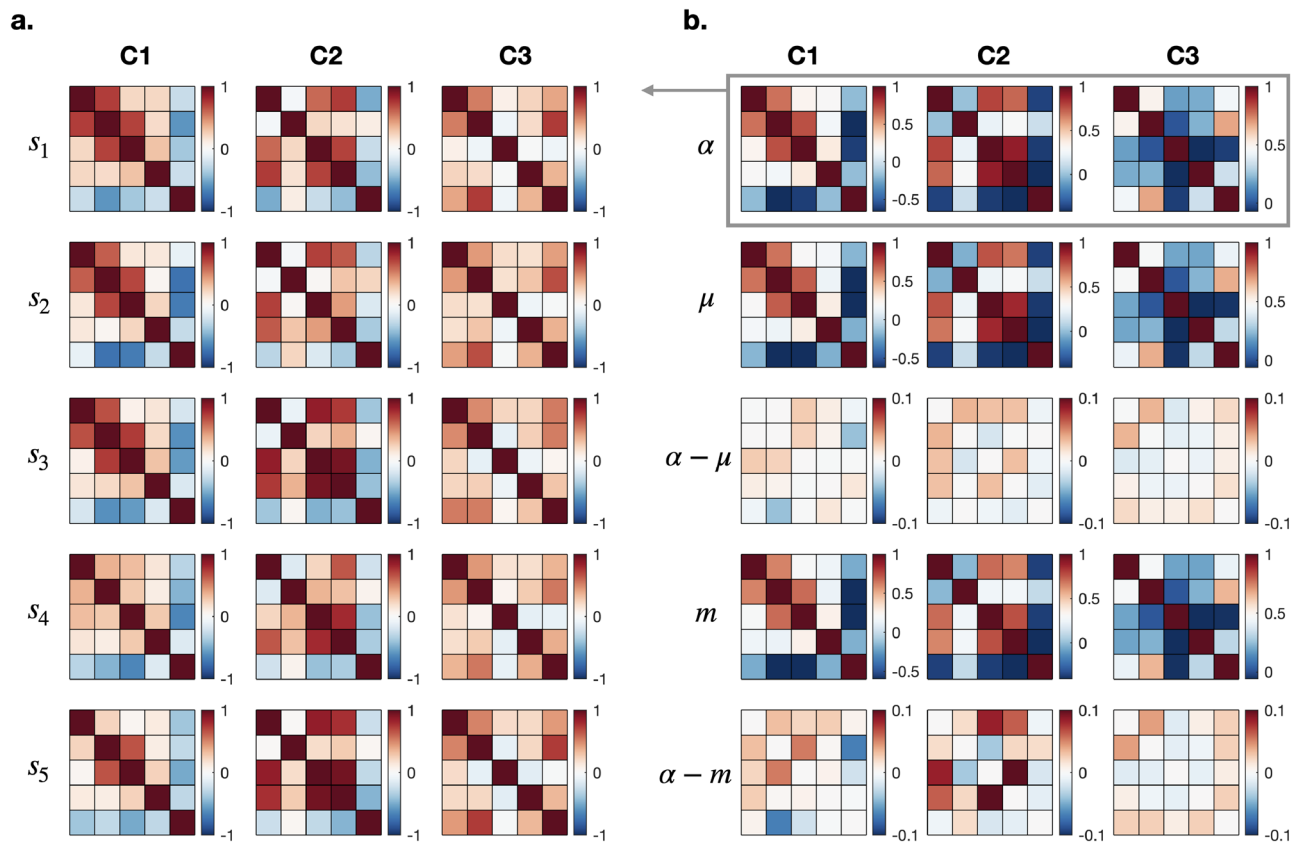


Figure 2. Comparison of the geometric mean (μ) and Euclidean average (m) for 30 samples generated from the three model correlation matrices (C1, C2, and C3) shown in the first row of (b), enclosed by a gray rectangle. Among the thirty samples slightly deviated from each model correlation matrix obtained by adding noise (the model matrix is denoted as α for C1, C2, and C3), five exemplary samples for each model correlation matrix are presented in (a). The third and fifth rows in (b) indicate the element-wise differences from the model correlation matrix (α). By visual analysis, greater differences are found in the Euclidean average compared to the geometric mean.

		$n = 5$	$n = 10$	$n = 15$	$n = 30$
$N = 20$	CORR	0.183	0.487	0.839	2.249
	SPD	0.008	0.016	0.035	0.073
$N = 50$	CORR	0.383	0.988	1.690	5.365
	SPD	0.017	0.032	0.069	0.171
$N = 100$	CORR	0.813	2.132	3.387	10.743
	SPD	0.035	0.066	0.145	0.341

Table 1. Average computational time (in s) to compute Fréchet mean of N correlation matrices of size ($n \times n$) under correlation and SPD geometry.

testing of two sets comprising random 30 samples (per set) derived from a group C_1 and the two different groups (i.e., C_1 and C_2), using Biswas–Ghosh (BG) and Wasserstein (WASS)-based tests.

Real data analysis: EEG motor imagery dataset. In this example, we show efficacy of our proposed framework on the EEG motor movement/imagery dataset⁵⁷, which is publicly available at PhysioNet⁵⁸. To briefly describe, the dataset is a collection of EEG recordings of one and two minutes duration from 109 volunteers. Subjects performed 14 experimental runs of different motor and imagery tasks whose neural activities were recorded with 64-channel EEG using the BCI2000 system⁵⁹. Among many tasks, we are interested in distinguishing the imagery operation of one or both fists from those of feet under the correlation-based functional network perspective.

To show validity of the proposed method, we took recordings from a randomly chosen single subject (S007) and extracted time series for tasks involving both fists and both feet. For preprocessing, we first selected 32 out

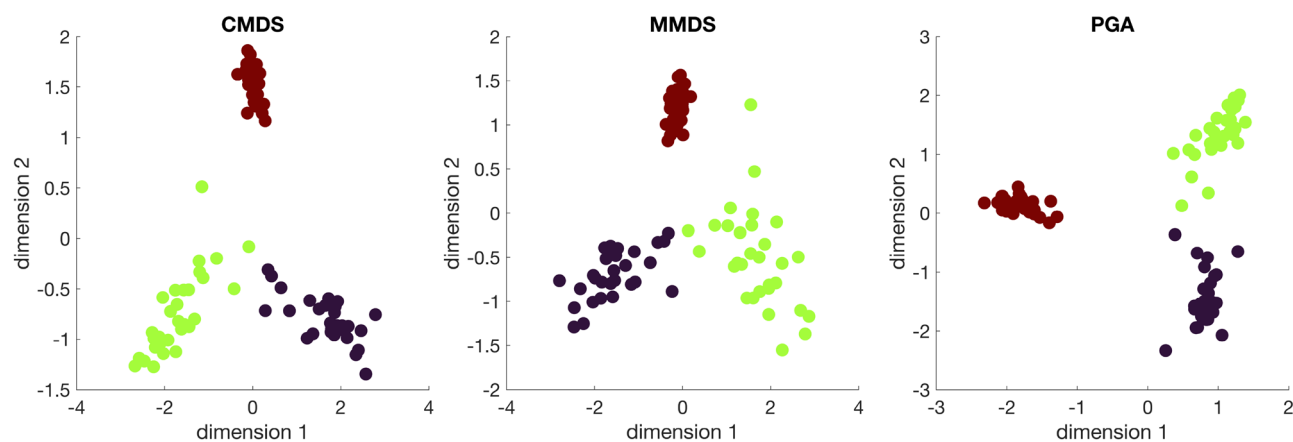


Figure 3. Embeddings of 90 correlation matrices from three model correlation matrices (C1, C2 and C3) by dimension reduction algorithms of classical multidimensional scaling (CMDS), metric multidimensional scaling (MMDS), and principal geodesic analysis (PGA).

	Silhouette			CH index		
	2	3	4	2	3	4
K	2	3	4	2	3	4
k -means	0.46	0.73	0.51	0.98	3.82	2.71
k -medoids	0.47	0.76	0.55	0.98	6.92	4.60
Spectral clustering	0.46	0.75	0.51	0.98	6.92	4.58

Table 2. Empirical cluster quality measures for cluster partitions obtained by proposed algorithms under varying number of clusters K . CH indices are scaled by 0.01 for demonstration purpose.

of 64 channels (Fc5, Fc1, Fc2, Fc6, C3, Cz, C4, Cp5, Cp1, Cp2, Cp6, Fpz, Af7, Afz, Af8, F5, F1, F2, F6, Ft7, T7, T9, Tp7, P7, P3, Pz, P4, P8, Po3, Po4, O1, O2), removing the other 32 channels as they are marked as bad channels whose signals are either flat or show excessive signal-to-noise ratio⁵⁷. We applied a Butterworth IIR band-pass filter with cutoffs at 7 and 35Hz. For network-based analysis, we epoched the filtered signals from every stimulus onset to a second after the onset (161 samples). Pearson correlation coefficients were constructed from the epoched signals. This led to a total of 45 correlation matrices where 23 are for feet, and 22 are for fists. Each correlation matrix is a 32×32 matrix whose rows and columns correspond to 32 remaining channels. All 45 matrices were verified to be full-rank so that they are indeed proper objects for analysis on the correlation manifold.

We computed Fréchet means of two classes under different geometries as a preliminary step in exploratory data analysis. As shown in Fig. 5, it appears clear that all three geometries were somehow capable of capturing distinct patterns across different classes, which is an expected phenomenon considering the nature of the data. However, when we consider differences in two mean matrices, the correlation manifold identified local heterogeneity well compared to the other two geometries.

We also performed hypothesis testing of equal distributions on the two-class correlation data under different geometries. For both Biswas–Ghosh (BG) and Wasserstein (WASS)-based tests, we ran 10^5 permutations to guarantee the credibility of the resampling-based procedures. As summarized in Table 3, both tests showed significant empirical p values upon the correlation manifold structure at the significance level $\alpha = 5\%$, suggesting that two empirical measures are statistically distinguishable on the correlation manifold only. An interesting observation is that although the other two geometries did not indicate a significant difference between the two classes, the SPD geometry showed dominant results against the Euclidean space assumption, and its empirical p value by the Wasserstein-based test is close to the cutoff value. This result aligns with our previous observation in the sense that a Riemannian approach can be a proper alternative that reflects intrinsic nonlinearity on the space of functional connectivities⁵⁵. The statement, however, does not necessarily preclude absolute dominance of a specific geometry against the others.

Lastly, we demonstrated how a choice of manifold structure affects its low-dimensional embedding. In Fig. 6, we presented 2-dimensional embeddings from the algorithms we introduced in our paper under three different geometries. At a glance, all embeddings seem to show entangled patterns for two separate classes. However, the correlation manifold shows the largest degree of distinction between two classes across all algorithms, while the Euclidean assumption does little separation. Similar to the previous experiments with other tasks, it is worth mentioning that the SPD geometry locates between the other two geometries. This partially supports that the SPD geometry may still be employed on its submanifold of correlation matrices at the cost of performance and concerns in choosing a more suitable geometry.

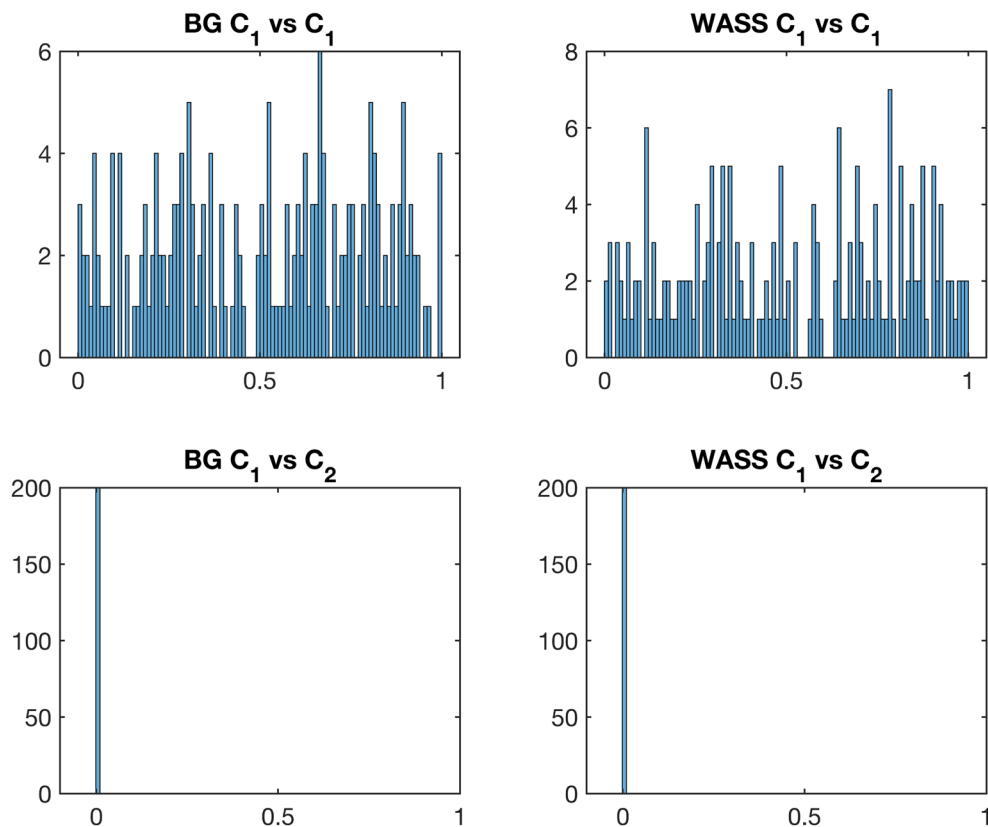


Figure 4. Hypothesis testing for two sets of correlation matrices driven from a model correlation matrix C_1 or from two different model correlation matrices C_1 and C_2 . Each set contains 30 random samples from the model correlation matrix. Empirical distributions of 200 p values by Biswas–Ghosh (BG) and Wasserstein (WASS)-based tests are presented.

Discussion

As a basic tool for functional network analysis, the correlation matrix contains more information as a whole than the sum of independent pairwise correlation coefficients. Thus, the correlation matrix may well be dealt with as manifold-valued objects with corresponding geometric structures. In recognition of the importance for operation over the proper manifold, a growing number of studies have analyzed the correlation matrix under the SPD geometry^{19–23}.

One critical limitation to considering correlation matrices simply as objects on SPD manifold is that operations with correlation matrices do not necessarily take the form of the correlation matrix and thus demand a post-hoc step to constrain unit diagonal elements. In our previous study⁵⁵, we iteratively normalized output matrices (derived from the correlation-matrix operation) into correlation matrices at each intermediate step. Although this heuristic works well in most cases as shown in our simulation study, this approach is not an exact solution. The current study resolves this naive solution by implementing operations over the correlation matrix manifold space known as *elliptope*²⁴, a mathematical space whose computational routines have been little known^{25,26}. Two recent seminal works^{27,29} make the construction of the correlation matrix space computationally feasible and lay the basis of the current paper.

Based on these recent developments in the quotient geometry of the correlation manifold, we numerically implemented computational operations over the correlation matrix space. We then presented the most fundamental analyses and inferential algorithms for practical use in functional network analysis, including measures of central tendency, cluster analysis, hypothesis testing, and low-dimensional embedding. The simulation result suggests the effectiveness of analysis on the correlation manifold. In our simulation, the SPD-based approach with post-hoc normalization shows comparable performance against the correlation-based approach. Nevertheless, the proposed framework is expected to be theoretically more sound. For example, our experiment, as shown in Fig. 2, revealed that even the most straightforward task of finding the mean over perturbed correlation matrices around the identity matrix benefits from the dedicated geometric structure. Our real data example demonstrated significance of our proposed approach in localizing differentiating inter-region connections between two classes. Furthermore, hypothesis testing and low-dimension embedding examples even provided grounds to argue that the correlation manifold was the only geometry that revealed difference in two clearly distinct classes of correlation matrices. It should be noted that the advantage of functional network analysis under the correlation-manifold framework may not always be conspicuous compared to the SPD manifold or Euclidean treatment in

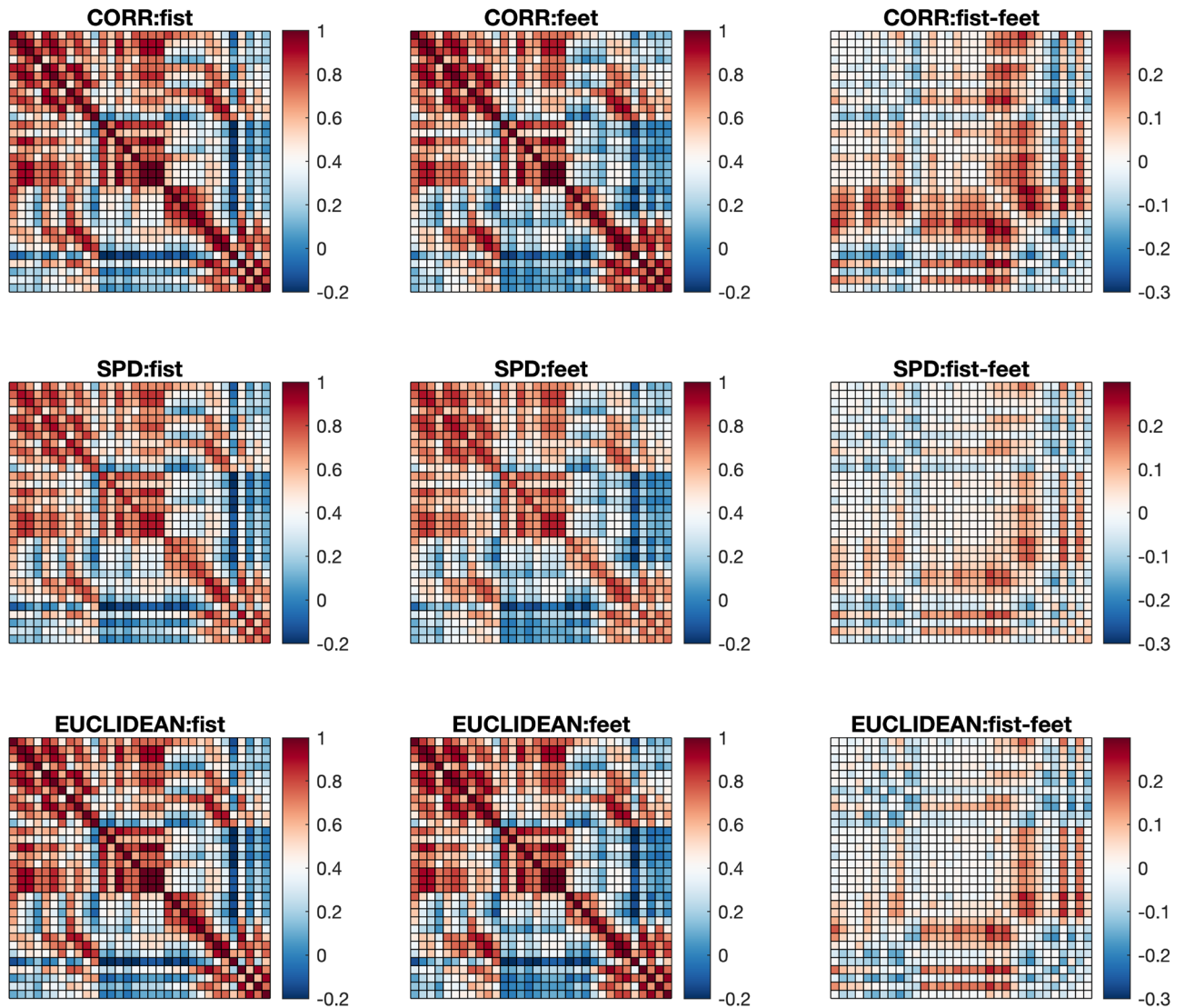


Figure 5. Fréchet mean of correlation matrices from fists and feet classes. Rows indicate a choice of geometric structure by viewing the matrices as elements of the correlation manifold (top), the SPD manifold (middle), and the Euclidean space (bottom). First two columns are Fréchet means for the fists and the feet classes, respectively. The last column is difference in Fréchet means of two classes.

Algorithms	Geometries		
	CORR	SPD	Euclidean
WASS	0.019*	0.056	0.108
BG	0.037*	0.095	0.144

Table 3. Empirical p values on two classes of correlation matrices obtained from the Wasserstein (WASS) and Biswas–Ghosh (BG)-based permutation tests under three geometries. Numeric values with an asterick * indicate the null hypothesis of equal distributions is rejected at the significance level of 5%.

machine learning or statistical analysis. However, it is mathematically more exact and consistent than the SPD framework in considering inter-dependence among edges in the functional brain network.

Despite its mathematical consistency and superb performance, there remains a number of issues to be addressed which set major directions for future studies. First, the correlation manifold structure provides an effective geometric framework at the cost of increased computational burden compared to that of the ambient SPD manifold. In our experiments, we observed that pertained computations were within a reasonable time range for a general network size of 30 (Table 1). Still, its nested iterative nature has its potential to hamper wider use

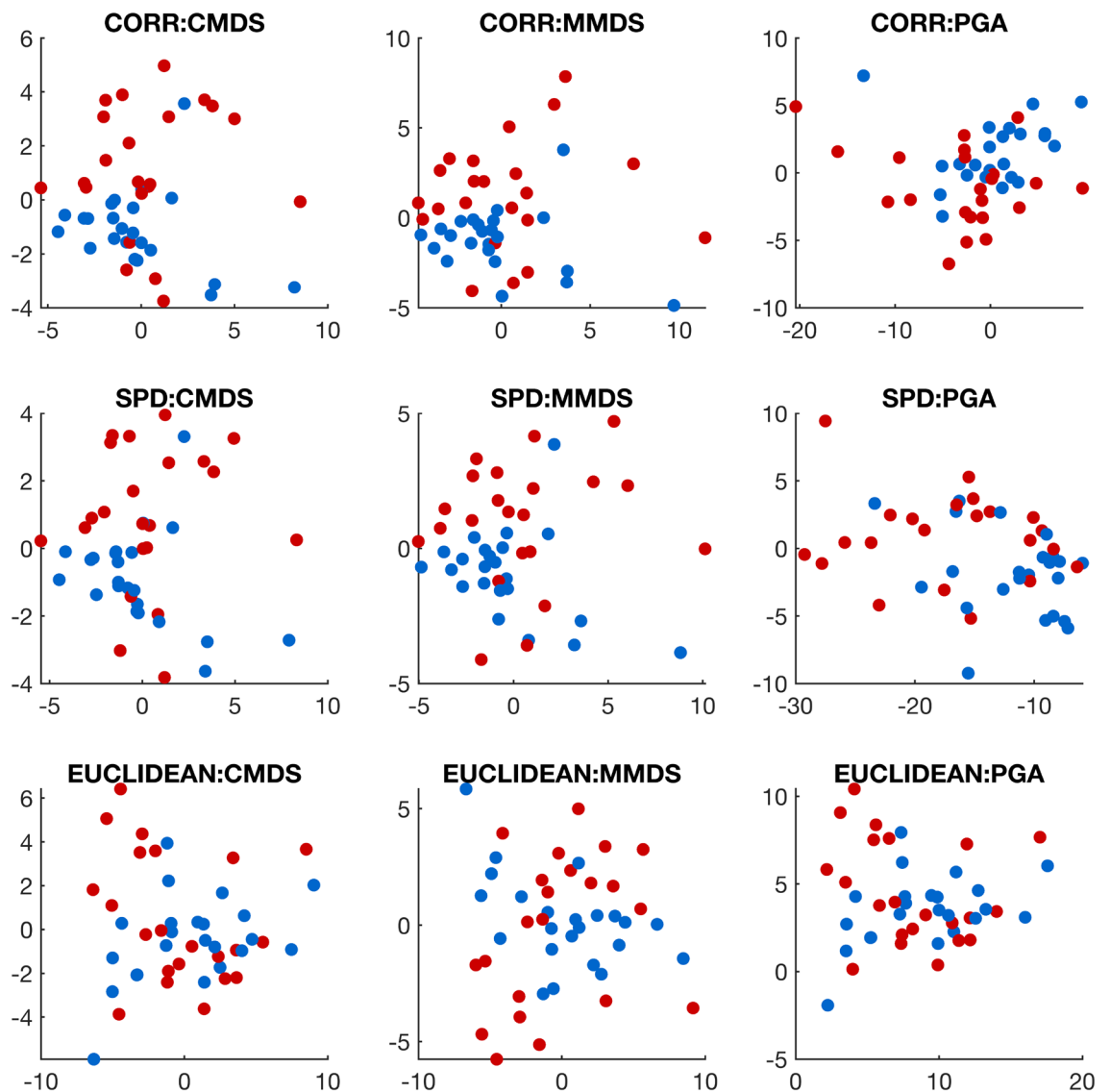


Figure 6. Embeddings onto 2-dimensional Euclidean space for a collection of motor imagery correlation matrices. Rows correspond to a choice of geometric structure by viewing the matrices as elements of the correlation manifold (top), the SPD manifold (middle), and the Euclidean space (bottom). Columns represent algorithms used under each geometry, including classical multidimensional scaling (left), metric multidimensional scaling (center), and principal geodesic analysis (right). Colors represent two classes of fists (in red) and feet (in blue).

in practice. This necessitates to devise a set of numerical routines that can dramatically reduce computational costs, which we view as an invaluable opportunity. The current correlation analysis methods also inherit the same practical limitation of the SPD in the neuroimaging analysis. When the number of scans for resting-state fMRI is small compared to the network size, correlation matrices are likely to be rank-deficient. In this case, one may consider estimating correlation matrices under assumptions like sparsity, strict non-singularity, and others^{60,61}.

Besides the functional network analysis, the correlation manifold would be applied to other neuroimaging research fields such as Representational Similarity Analysis (RSA). RSA is a multivariate technique that explores brain processes according to similarity matrices of the brain responses to the different types of stimuli^{62,63}. In the RSA, the similarity is generally defined in terms of the correlation matrix. Shahbazi et al.⁶⁴ showed more accurate quantification of representational similarities over the SPD manifold than the Euclidean operation. Considering the similarity definition based on the correlation matrix, the correlation manifold proposed in the current study would be a more appropriate choice for RSA.

The application of the correlation manifold is not restricted to brain research. It extends to any other research areas that utilize correlation matrices. For example, correlation matrix is a popular form of data in financial markets^{65–67}. The geometric approach to correlation-based functional network analysis is still nascent in the neuroimaging and other related research communities. We wrapped all the algorithms in the current paper as a toolbox of MATLAB (Mathworks, Inc. USA) called CORRbox, which is freely available on GitHub (<https://github.com>).

com/kisungyou/papers). We expect that CORRbox would be of great use in analyzing matrix representations of functional networks by taking advantage of exact representations and operations over the proper manifold of the correlation matrices.

Data availability

The CORRbox, a MATLAB toolbox for learning with data on the correlation manifold, is publicly available at the github repository (<https://github.com/kisungyou/papers>) along with examples.

Received: 8 June 2022; Accepted: 27 September 2022

Published online: 22 October 2022

References

- Park, H.-J. & Friston, K. Structural and functional brain networks: From connections to cognition. *Science* **342**, 1238411. <https://doi.org/10.1126/science.1238411> (2013).
- Biswal, B., Zerrin Yetkin, F., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med.* **34**, 537–541. <https://doi.org/10.1002/mrm.1910340409> (1995).
- Brookes, M. J. *et al.* Measuring functional connectivity using MEG: Methodology and comparison with fMRI. *Neuroimage* **56**, 1082–1104. <https://doi.org/10.1016/j.neuroimage.2011.02.054> (2011).
- Cohen, M. X. *Analyzing Neural Time Series Data: Theory and Practice. Issues in Clinical and Cognitive Neuropsychology* (The MIT Press, 2014).
- Yahata, N. *et al.* A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. Commun.* **7**, 11254. <https://doi.org/10.1038/ncomms11254> (2016).
- Lee, D. *et al.* Analysis of structure–function network decoupling in the brain systems of spastic diplegic cerebral palsy. *Hum. Brain Mapp.* **38**, 5292–5306. <https://doi.org/10.1002/hbm.23738> (2017).
- Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38. <https://doi.org/10.1038/nm.4246> (2017).
- Finn, E. S. *et al.* Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671. <https://doi.org/10.1038/nn.4135> (2015).
- Jang, C. *et al.* Individuality manifests in the dynamic reconfiguration of large-scale brain networks during movie viewing. *Sci. Rep.* **7**, 41414. <https://doi.org/10.1038/srep41414> (2017).
- Calhoun, V. D., Miller, R., Pearlson, G. & Adalı, T. The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* **84**, 262–274. <https://doi.org/10.1016/j.neuron.2014.10.015> (2014).
- Allen, E. A. *et al.* Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24**, 663–676. <https://doi.org/10.1093/cercor/bhs352> (2014).
- Monti, R. P. *et al.* Estimating time-varying brain connectivity networks from functional MRI time series. *Neuroimage* **103**, 427–443. <https://doi.org/10.1016/j.neuroimage.2014.07.033> (2014).
- Jeong, S.-O., Pae, C. & Park, H.-J. Connectivity-based change point detection for large-size functional networks. *Neuroimage* **143**, 353–363. <https://doi.org/10.1016/j.neuroimage.2016.09.019> (2016).
- Preti, M. G., Bolton, T. A. & Van De Ville, D. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage* **160**, 41–54. <https://doi.org/10.1016/j.neuroimage.2016.12.061> (2017).
- Dosenbach, N. U. F. *et al.* Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361. <https://doi.org/10.1126/science.1194144> (2010).
- Siman-Tov, T. *et al.* Early age-related functional connectivity decline in high-order cognitive networks. *Front. Aging Neurosci.* <https://doi.org/10.3389/fnagi.2016.00330> (2017).
- Leonardi, N. *et al.* Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *Neuroimage* **83**, 937–950. <https://doi.org/10.1016/j.neuroimage.2013.07.019> (2013).
- Park, B., Kim, D.-S. & Park, H.-J. Graph independent component analysis reveals repertoires of intrinsic network components in the human brain. *PLoS ONE* **9**, e82873. <https://doi.org/10.1371/journal.pone.0082873> (2014).
- Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P. & Thirion, B. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010*, Vol. 6361 200–208 (Springer, 2010). https://doi.org/10.1007/978-3-642-15705-9_25.
- Yamin, A. *et al.* Comparison of brain connectomes using geodesic distance on manifold: A twins study. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 1797–1800 (IEEE, Venice, Italy, 2019). <https://doi.org/10.1109/ISBI.2019.8759407>.
- Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. & Kolaczyk, E. D. Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* <https://doi.org/10.1214/16-AOAS1015> (2017).
- Deligianni, F. *et al.* A probabilistic framework to infer brain functional connectivity from anatomical connections. In *Information Processing in Medical Imaging* Vol. 6801 (eds Székely, G. & Hahn, H. K.) 296–307 (Springer, 2011). https://doi.org/10.1007/978-3-642-22092-0_25.
- Abbas, K. *et al.* Geodesic distance on optimally regularized functional connectomes uncovers individual fingerprints. *Brain Connect.* **11**, 333–348. <https://doi.org/10.1089/brain.2020.0881> (2021).
- Tropp, J. A. Simplicial faces of the set of correlation matrices. *Discrete Comput. Geom.* **60**, 512–529. <https://doi.org/10.1007/s00454-017-9961-0> (2018).
- Grubišić, I. & Pietersz, R. Efficient rank reduction of correlation matrices. *Linear Algebra Appl.* **422**, 629–653. <https://doi.org/10.1016/j.laa.2006.11.024> (2007).
- Nielsen, F. & Sun, K. Clustering in Hilbert’s projective geometry: The case studies of the probability simplex and the ellipsope of correlation matrices. In *Geometric Structures of Information* (ed. Nielsen, F.) 297–331 (Springer, 2019). https://doi.org/10.1007/978-3-030-02520-5_11.
- David, P. A *Riemannian Quotient Structure for Correlation Matrices with Applications to Data Science*. PhD Thesis, Claremont Graduate University (2019).
- Pennec, X. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.* **25**, 127–154. <https://doi.org/10.1007/s10851-006-6228-4> (2006).
- Thanwerdas, Y. & Pennec, X. Geodesics and curvature of the quotient-affine metrics on full-rank correlation matrices. In *Geometric Science of Information* Vol. 12829 (eds Nielsen, F. & Barbaresco, F.) 93–102 (Springer, 2021). https://doi.org/10.1007/978-3-030-80209-7_11.
- Hall, B. C. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. No. 222 in *Graduate Texts in Mathematics* 2nd edn. (Springer, 2015).

31. Grove, K. & Karcher, H. How to conjugate C 1-close group actions. *Math. Z.* **132**, 11–20. <https://doi.org/10.1007/BF01214029> (1973).
32. Afsari, B. Riemannian L^1 center of mass: Existence, uniqueness, and convexity. *Proc. Am. Math. Soc.* **139**, 655–655. <https://doi.org/10.1090/S0002-9939-2010-10541-5> (2011).
33. Arnaudon, M., Barbaresco, F. & Yang, L. Medians and means in Riemannian geometry: Existence, uniqueness and computation. In *Matrix Information Geometry* (eds Nielsen, F. & Bhatia, R.) 169–197 (Springer, 2013). https://doi.org/10.1007/978-3-642-30232-9_8.
34. Weiszfeld, E. Sur le point pour lequel la Somme des distances de n points donnees est minimum. *Tohoku Math. J. First Ser.* **43**, 355–386 (1937).
35. Weiszfeld, E. & Plastria, F. On the point for which the sum of the distances to n given points is minimum. *Ann. Oper. Res.* **167**, 7–41. <https://doi.org/10.1007/s10479-008-0352-z> (2009).
36. Fletcher, P. T., Venkatasubramanian, S. & Joshi, S. The geometric median on Riemannian manifolds with application to robust atlas estimation. *Neuroimage* **45**, S143–S152. <https://doi.org/10.1016/j.neuroimage.2008.10.052> (2009).
37. Aftab, K., Hartley, R. & Trumpp, J. Generalized weiszfeld algorithms for Lq optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 728–745. <https://doi.org/10.1109/TPAMI.2014.2353625> (2015).
38. MacQueen, J. B. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* Vol. 1 (eds Cam, L. M. L. & Neyman, J.) 281–297 (University of California Press, 1967).
39. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137. <https://doi.org/10.1109/TIT.1982.1056489> (1982).
40. Kaufman, L. & Rousseeuw, P. J. Partitioning around medoids (Program PAM). In *Wiley Series in Probability and Statistics* 68–125 (Wiley, 1990). <https://doi.org/10.1002/9780470316801.ch2>.
41. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. *Wiley Series in Probability and Mathematical Statistics* (Wiley, 2005).
42. von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416. <https://doi.org/10.1007/s11222-007-9033-z> (2007).
43. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905. <https://doi.org/10.1109/34.868688> (2000).
44. Ng, A., Jordan, M. & Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* (eds Dietterich, T. G., Becker, S. & Ghahramani, Z.), Vol. 14 (MIT Press, 2002).
45. Chung, F. R. K. *Spectral Graph Theory*. No. 92 in *Regional Conference Series in Mathematics* (Published for the Conference Board of the Mathematical Sciences. American Mathematical Society, Providence, RI, 1997).
46. Zelnik-manor, L. & Perona, P. Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems* (eds Saul, L., Weiss, Y. & Bottou, L.), Vol. 17 1601–1608 (MIT Press, 2004).
47. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M. & Perona, I. An extensive comparative study of cluster validity indices. *Pattern Recogn.* **46**, 243–256. <https://doi.org/10.1016/j.patcog.2012.07.021> (2013).
48. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7) (1987).
49. Calinski, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. Theory Methods* **3**, 1–27. <https://doi.org/10.1080/03610927408827101> (1974).
50. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification* 2nd edn. (Wiley, 2001).
51. Borg, I. & Groenen, P. J. F. *Modern Multidimensional Scaling: Theory and Applications* (*Springer Series in Statistics*) (Springer, 1997).
52. Fletcher, P., Lu, C., Pizer, S. & Joshi, S. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging* **23**, 995–1005. <https://doi.org/10.1109/TMI.2004.831793> (2004).
53. de Leeuw, J. Applications of convex analysis to multidimensional scaling. In *Recent Developments in Statistics* (eds Barra, J. et al.) 133–146 (North Holland Publishing Company, 1977).
54. Pearson, K. L. I. I. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559–572. <https://doi.org/10.1080/14786440109462720> (1901).
55. You, K. & Park, H.-J. Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity. *Neuroimage* **225**, 117464. <https://doi.org/10.1016/j.neuroimage.2020.117464> (2021).
56. Lehmann, E. L. & Romano, J. P. *Testing Statistical Hypotheses*. *Springer Texts in Statistics* 3rd edn. (Springer, 2005).
57. Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N. & Wolpaw, J. R. EEG Motor Movement/Imagery Dataset. <https://doi.org/10.13026/C28G6P> (2009).
58. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* <https://doi.org/10.1161/01.CIR.101.23.e215> (2000).
59. Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N. & Wolpaw, J. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* **51**, 1034–1043. <https://doi.org/10.1109/TBME.2004.827072> (2004).
60. Fan, J., Liao, Y. & Liu, H. An overview of the estimation of large covariance and precision matrices. *Economet. J.* **19**, C1–C32. <https://doi.org/10.1111/ectj.12061> (2016).
61. Lam, C. High-dimensional covariance matrix estimation. *WIREs Comput. Stat.* <https://doi.org/10.1002/wics.1485> (2020).
62. Kriegeskorte, N. Representational similarity analysis—Connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* <https://doi.org/10.3389/neuro.06.004.2008> (2008).
63. Kriegeskorte, N. & Kievit, R. A. Representational geometry: Integrating cognition, computation, and the brain. *Trends Cogn. Sci.* **17**, 401–412. <https://doi.org/10.1016/j.tics.2013.06.007> (2013).
64. Shahbazi, M., Shirali, A., Aghajan, H. & Nili, H. Using distance on the Riemannian manifold to compare representations in brain and in models. *Neuroimage* **239**, 118271. <https://doi.org/10.1016/j.neuroimage.2021.118271> (2021).
65. Mantegna, R. Hierarchical structure in financial markets. *Eur. Phys. J. B* **11**, 193–197. <https://doi.org/10.1007/s100510050929> (1999).
66. Bonanno, G., Caldarelli, G., Lillo, F. & Mantegna, R. N. Topology of correlation-based minimal spanning trees in real and model markets. *Phys. Rev. E* **68**, 046130. <https://doi.org/10.1103/PhysRevE.68.046130> (2003).
67. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J. & Kanto, A. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E* **68**, 056110. <https://doi.org/10.1103/PhysRevE.68.056110> (2003).

Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (2022M3E5E8018285).

Author contributions

K.Y. conceptualized and developed methodology and software. H.J.P. initiated the project, received a grant and conceptualized the method. K.Y. and H.J.P. wrote the main manuscript text. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21376-0>.

Correspondence and requests for materials should be addressed to H.-J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022