HHS Public Access

Author manuscript

IEEE J Biomed Health Inform. Author manuscript; available in PMC 2023 September 15.

Published in final edited form as:

IEEE J Biomed Health Inform. 2023 May ; 27(5): 2411–2422. doi:10.1109/JBHI.2023.3250711.

Estimating Outlier-Immunized Common Harmonic Waves for Brain Network Analyses on the Stiefel Manifold

Hongmin Cai [Senior Member, IEEE],
School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China

Huan Liu,
School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China

Defu Yang [Member, IEEE],
Department of Psychiatry and Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

Guorong Wu,
Department of Psychiatry and Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

Bin Hu [Fellow, IEEE],
School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China

Jiazhou Chen [Member, IEEE]
School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China

Abstract

Since brain network organization is essentially governed by the harmonic waves derived from the Eigen-system of the underlying Laplacian matrix, discovering the harmonic-based alterations provides a new window to understand the pathogenic mechanism of Alzheimer's disease (AD) in a unified reference space. However, current reference (common harmonic waves) estimation studies over the individual harmonic waves are often sensitive to outliers, which are obtained by averaging the heterogenous individual brain networks. To address this challenge, we propose a novel manifold learning approach to identify a set of outlier-immunized common harmonic waves. The backbone of our framework is calculating the geometric median of all individual harmonic waves on the Stiefel manifold, instead of Fréchet mean, thus improving the robustness of learned common harmonic waves to the outliers. A manifold optimization scheme with theoretically guaranteed convergence is tailored to solve our method. The experimental results on synthetic data and real data demonstrate that the common harmonic waves learned by our approach are not only more robust to the outliers than the state-of-the-art methods, but also provide a putative imaging biomarker to predict the early stage of AD.

Corresponding author: Jiazhou Chen. 09chenjiazhou@gmail.com.

Keywords

Alzheimer's disease; brain network; harmonic waves; manifold optimization

I. Introduction

ALZHEIMER'S disease (AD), characterized by memory loss and cognitive abnormalities, is one of the most common forms of dementia [1], [2]. The World Alzheimer Report 2018 [3] reported that there are 50 million people all the world lived with dementia. As a result, the sustainable development of society is affected. More importantly, the neuropathological mechanism of AD is still unknown, and no treatment has been developed to reverse or stop the progression of AD [4], [5]. Therefore, an increasing number of studies have turned their attention to predicting AD at an early stage so that appropriate interventions can be used to slow down its progression [6]. Mild cognitive impairment (MCI) is a mid-prodromal period of memory impairment that is often considered a transitional state from a healthy individual to an individual with AD [7]. The accurate identification of MCI patients who will progress to AD is essential to exert possible therapeutic interventions to delay this progression. However, the early prediction of AD remains a difficult challenge, since MCI exhibits substantial individual heterogeneity and mild symptoms.

Abnormal connectivity between distinct brain regions manifests much earlier than the emergence of the earliest symptoms of AD; thus, accurately identifying brain network changes facilitates the early prediction of AD [8]. The rapid development of noninvasive neuroimaging and neurophysiological techniques allows us to capture multimodal brain images from the same samples, providing an efficient, feasible, and noninvasive way to investigate structural brain connectivity *in vivo*, which is consisted of the number of white matter fiber connections between underlying regions of interest (ROIs) [9], [10].

Due to the high dimensionality of brain connectome data, it is a common practice to analyze node-wise graph variables such as local clustering coefficient, centrality, efficiency, modularity and small-worldness, instead of using whole-brain connectivity information. For instance, He et al. [11] found increased clustering coefficient and shortest paths in AD, implying an abnormal small-world property; Yao et al. [12] found the greatest clustering coefficient and the longest absolute path length in AD; Tijms et al. [13] found decreased small-world index in AD; Pereira et al. [14] found that all patient groups exhibited increased path length, reduced transitivity, and increased modularity, and the patient group showed decreased small-world index. By doing so, however, it becomes difficult to discover topological patterns which is an essential aspect of network analyses [15]. Furthermore, brain network changes have been proven to be correlated with the underlying structural pathology [16]. In addition, the unique propagation patterns in existing large-scale brain networks mainly drive the abnormal deposition of AD pathologies [17], [18], [19], [20]. There are an increasing number of brain network analyses focusing on investigating the attributes of brain networks [21], [22], [23], [24], [25].

Recently, many studies have applied the graph Laplacian operator to the adjacency matrix of each individual brain for the purpose of studying the spreading process of neuropathological

events of neurodegenerative diseases [26], [27], [28]. The eigensystem of the underlying Laplacian matrix, which is referred to as harmonic waves in this paper, reveals the propagation property of the brain network [29], [30], [31] (shown in the left part of Fig. 1). Combined with the orthogonality of harmonic waves, it is possible to distinguish different clinical cohorts by examining alterations based on harmonic waves. In summary, common harmonic waves obtained by unifying individual eigensystems are required to achieve classification. Although treating the high-dimensional brain networks as regular matrices and applying Euclidean algebra to them to obtain the average brain network is straightforward, such an approach destroys the essential network topological structure of the resulting group-mean brain network (demonstrated in the upper right of Fig. 1). Since each eigensystem of the brain network is an orthogonal matrix, which is considered to be a point on the Stiefel manifold, it provides the possibility to estimate the common harmonic waves by finding the manifold center of all those individual eigensystems on the Stiefel manifold, as illustrated in the lower part in Fig. 1. To this end, the manifold analysis methods are proposed to solve the eigensystem on the manifold. For example, Huang et al. [32] proposed polynomial expansion method of the Laplace-Beltrami operator to obtain the eigensystem and use it to solve heat diffusion on a manifold. Chen et al. [33] proposed a manifold-based harmonic network analysis approach to identify a set of region-adaptive harmonic wavelets that represent the common network topology across individuals.

However, the methods mentioned above ignore the influence of outliers. The outliers in brain network data will severely affect the estimation of manifold centrality. A straightforward solution is outlier deletion; however, this approach is not appropriate for the analysis of a few brain network samples [34]. In this regard, Chen et al. [35] proposed the estimation of the common harmonic waves by calculating the Fréchet mean, where each single harmonic wave represents a unique brain network neuropathological burden spreading pattern. Unfortunately, the Fréchet mean is sensitive to outliers since any point located on the manifold is dragged to infinity without bound. Therefore, designing an outlier-immunized center estimator to identify the common harmonic waves on a Stiefel manifold remains unsolved.

To address this challenge, instead of using a rigid averaging operation on Euclidean space, as demonstrated in Fig. 1, we propose a novel manifold harmonic learning approach to derive outlier-immunized common harmonic waves for each individual wave on a Stiefel manifold. Our method aims to calculate the geometric median of the population harmonic waves, which is formulated as an optimization problem to find manifold centers by minimizing the geodesic distance across all individual waves on the Stiefel manifold. Numerical schemes are designed to iteratively solve the minimization problem on the manifold, and its convergence is theoretically proven. We show that the geometric median is a robust estimator of centrality and thus avoids outlier corruption. After obtaining the common harmonic waves of the network population, we assess its statistical performance on both synthetic data and ADNI neuroimaging data. Extensive experiments on synthetic data have demonstrated that our proposed method achieves superior identification accuracy in estimating the common harmonic waves containing an increased number of brain network outliers. For real data, our manifold learning approach achieves higher discriminatory power than the other comparative methods in distinguishing cognitively normal (CN),

early-stage mild cognitive impairment (EMCI), and late-stage mild cognitive impairment (LMCI). Finally, we identify a set of neuroimaging biomarkers derived from the significant common harmonic waves that are highly associated with the mini-mental state examination (MMSE) score, a clinical test to measure cognitive impairment. Our study provides three main contributions:

1. We develop a novel manifold learning approach to estimate outlier-immunized common harmonic waves by calculating the geometric median, which shows better robustness to outliers than arithmetic mean or Fréchet mean in constructing the group mean.
2. A specialized manifold optimization algorithm for our proposed method is designed to ensure that the optimal numerical solution is obtained. Its theoretical proof of convergence is elaborated.
3. Extensive experiments demonstrate that our method not only achieves higher diagnostic performance than other baseline methods but also identifies a new neuroimaging biomarker to predict the risk of progressing AD at the preclinical phase.

The organization of this article is as follows. The proposed method and its theoretical convergent-guaranteed numerical scheme are introduced in Section II. Extensive experiments on synthetic data and ADNI neuroimaging data are presented in Section III. A further discussion of the proposed method and experimental results is proposed in Section IV. The conclusion is given in Section V.

II. Method

The problem statement and mathematical modeling for estimating the outlier-immunized common harmonic waves is first explained in Section II-A. Then, the common harmonic wave optimization scheme is presented in Section II-B. Finally, the application of harmonic wave analysis is given in Section II-C. The notation is summarized in Table I.

A. Problem Statement and Mathematical Modeling

Generally, we adopt a graph representation $\mathcal{G} = (V, \varepsilon, \boldsymbol{W})$ to encode a complex brain network, where $V = \{v_i \mid i \in 1, \ldots N\}$ is the node set, and N stands for the number of nodes. Edge set $\varepsilon = \{e_{ij} \mid v_i, v_j\}$ represents the set of all possible connections between nodes. The weighted adjacency matrix $\boldsymbol{W} = [w_{i,j}]_{N \times N} \in \mathbb{R}^{N \times N}$ is used to store the connection strength between node i and node j.

The individual harmonic waves $\boldsymbol{\Phi} \in \mathbb{R}^{N \times P}$, which is the eigensystem of the Laplacian matrix, can be solved by the following minimization [35]:

$$\min_{\boldsymbol{\Phi} \in \mathbb{R}^{N \times P}} Tr\left(\boldsymbol{\Phi}^T \boldsymbol{L} \boldsymbol{\Phi}\right) s.t. \boldsymbol{\Phi}^T \boldsymbol{\Phi} = \boldsymbol{I}_P \tag{1}$$

where $Tr(\cdot)$ stands for the trace norm and $\boldsymbol{I}_P \in \mathbb{R}^{P \times P}$ is the identity matrix. A symmetry graph Laplacian matrix can be obtained by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ is the diagonal matrix. Each diagonal element $\boldsymbol{D}(i,i) = \sum_{j=1}^{N} w_{ij}$ in matrix $\boldsymbol{D}$ is represented as the summation of the connection strength between node i and its connected nodes. Note that the individual harmonic waves $\boldsymbol{\Phi} = \{\varphi_1, \varphi_2, \ldots, \varphi_P\}$ are sorted column by column from low to high frequency $\{\lambda_p \mid p = 1, \ldots, P, \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_P\}$ [36], showing that as the eigenvalue increases, the harmonic wave exhibits faster oscillatory patterns. We only reserve the top $P(P < N)$ harmonic waves for each brain network, since the higher frequency harmonic waves tend to be sensitive to potential noise. Assuming that a set of individual harmonic waves $\boldsymbol{\Phi}$ is an orthogonal matrix of $N \times P$, $\boldsymbol{\Phi}$ derived from a single brain network is reasonable to be treated as an instance on the Stiefel manifold [37].

Given the m brain network $\mathcal{G}_s(s = 1, \ldots, m)$, the individual harmonic waves $\boldsymbol{\Phi}_s$ can be optimized based on the energy function in (1). Our goal is to estimate the common harmonic waves $\boldsymbol{\Psi}$ by unifying the individual eigensystems from individual harmonic waves $\{\boldsymbol{\Phi}_s \mid s = 1, \ldots, m\}$. Specifically, we estimated the common harmonics by finding the Fréchet mean, which has the shortest l_2-norm geodesic distances to all observation points residing on the Stiefel manifold. Since Fréchet mean has been shown to be susceptible to outliers [34], we adopted a geometric median estimator to address this impact. This estimator can be used to effectively reduce the influence of outliers by calculating the l_1-norm geodesic distance instead of the l_2-norm geodesic distances. Therefore, the geometric median can be obtained by minimizing the summation of the l_1-norm geodesic distances to all individual harmonic waves $\{\boldsymbol{\Phi}_s\}$,

$$\arg\min_{\boldsymbol{\Psi}} \sum_{s=1}^{m} d(\boldsymbol{\Phi}_s, \boldsymbol{\Psi}) = \arg\min_{\boldsymbol{\Psi}} \sum_{s=1}^{m} \sqrt{P - Tr(\boldsymbol{\Phi}_s^T \boldsymbol{\Psi})} \tag{2}$$

where $d(\cdot)$ represents the l_1-norm geodesic distance, which is approximated by $d(\boldsymbol{\Phi}_s, \boldsymbol{\Psi}) = \sqrt{P - Tr(\boldsymbol{\Phi}_s^T \boldsymbol{\Psi})}$ [35]. As a result, by combining (1) and (2), the objective function is formulated as:

$$\begin{aligned} &\min_{\{\boldsymbol{\Phi}_s\}, \boldsymbol{\Psi}} \sum_{s=1}^{m} \left\{ Tr(\boldsymbol{\Phi}_s^T \boldsymbol{L}_s \boldsymbol{\Psi}_s) + \beta \sqrt{P - Tr(\boldsymbol{\Phi}_s^T \boldsymbol{\Psi})} \right\} \\ &\text{s.t. } \forall s : \boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s = \boldsymbol{I}_P \end{aligned} \tag{3}$$

where β is a scalar balancing of two terms in (3). The first term is used to ensure that each individual harmonic wave $\boldsymbol{\Phi}_s$ retains the topological structure of its own brain network. The second term ensures that the estimated common harmonic waves $\boldsymbol{\Psi}$ have the shortest geodesic distance summation to all individual harmonic waves $\{\boldsymbol{\Phi}_s\}$. The orthogonal constraint term can guarantee that all adjusted individual harmonic wave $\{\boldsymbol{\Phi}_s\}$ is still located on the Stiefel manifold.

B. Numerical Scheme

We propose using the gradient descent method under the alternating direction method of multipliers (ADMM) framework on the manifold [35] to solve the problem. Rewriting the constrained problem as an unconstrained minimization problem, the Lagrangian function is obtained as follows:

$$\min_{\{\boldsymbol{\Phi}_s\},\boldsymbol{\Psi}} \sum_{S=1}^{m} \boldsymbol{F}_{\boldsymbol{\Phi}_s,\boldsymbol{\Psi}} = \min_{\{\boldsymbol{\Phi}_s\},\boldsymbol{\Psi}} \sum_{S=1}^{m} \left\{ Tr\left(\boldsymbol{\Phi}_s^T \boldsymbol{L}_s \boldsymbol{\Phi}_s\right) + \beta\sqrt{P - Tr\left(\boldsymbol{\Phi}_s^T \boldsymbol{\Psi}\right)} + Tr\left(\boldsymbol{\Lambda}_s^T \left(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s - \boldsymbol{I}_P\right)\right) \right\} \tag{4}$$

where $\boldsymbol{\Lambda}_s$ is the Lagrangian matrix.

We solve the Lagrangian function in (4) by the following two steps.

Step 1: Optimize all individual harmonic waves $\{\Phi_s\}$ by fixing the common harmonic waves Ψ: Since each individual harmonic wave $\boldsymbol{\Phi}_s$ is independent, we can optimize each individual harmonic wave $\boldsymbol{\Phi}_s$ separately by simplifying (4) as:

$$\min_{\boldsymbol{\Phi}_s} \boldsymbol{F}_{\boldsymbol{\Phi}_s} = \min_{\boldsymbol{\Phi}_s} \{ Tr\left(\boldsymbol{\Phi}_s^T \boldsymbol{L}_s \boldsymbol{\Psi}_s\right) + \beta\sqrt{P - Tr\left(\boldsymbol{\Phi}_s^T \boldsymbol{\Psi}\right)} + Tr\left(\boldsymbol{\Lambda}_s^T \left(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s - \boldsymbol{I}_P\right)\right) \} \tag{5}$$

By setting the gradient $\nabla \boldsymbol{F}$ to zero, we obtain:

$$\frac{\partial \boldsymbol{F}_{\boldsymbol{\Phi}_s}}{\partial \boldsymbol{\Phi}_s} = 2\boldsymbol{L}_s \boldsymbol{\Phi}_s - \frac{\beta \boldsymbol{I}_N \boldsymbol{\Psi}}{2\sqrt{\left(P - Tr\left(\boldsymbol{\Phi}_s^T \boldsymbol{\Psi}\right)\right)}} + 2\boldsymbol{\Phi}_s \boldsymbol{\Lambda}_s^T = 0 \tag{6}$$

Since the objective function (5) is nonconvex and the Karush Kuhn Tucker (KKT) condition (6) with the square root term as the denominator is hard to solve, we adopt the following method [38], [39] to solve (5).

By introducing a specific Lagrangian multiplier θ for the third item, we obtain a partial Lagrangian relaxation (PLR, $f_\theta(\boldsymbol{\Phi}_s)$) form for the energy function (5):

$$\begin{aligned} &\min_{\boldsymbol{\Phi}_s} \{ Tr\left(\boldsymbol{L}_s \boldsymbol{\Phi}_s \boldsymbol{\Phi}_s^T\right) \\ &\quad + \beta\sqrt{Tr\left(\boldsymbol{I}_P - \boldsymbol{\Phi}_s^T \boldsymbol{\Psi}\right)} + \theta\left(P - Tr\left(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s\right)\right) \} \\ &\text{s.t.} \begin{pmatrix} \boldsymbol{I}_N & \boldsymbol{\Phi}_s \\ \boldsymbol{\Phi}_s^T & \boldsymbol{I}_P \end{pmatrix} \succeq 0 \end{aligned} \tag{7}$$

By choosing appropriate Lagrangian multiplier θ [40], $f_\theta(\boldsymbol{\Phi}_s)$ is a convex conic program problem that can be easily solved by the Frank-Wolfe algorithm [41], [42]. The procedure of optimizing individual harmonic waves $\boldsymbol{\Phi}_s$ includes the following four steps:

1. Initialize the parameters $i = 1$, $H = 100$.
2. Update $\theta_i \leftarrow \theta_{\min} + (i-1)\frac{\theta_{\max} - \theta_{\min}}{H}$, $i = 1, \ldots, H$
3. Apply the Frank-Wolfe algorithm to obtain an optimal solution $\mathbf{\Phi}_s^*$ of the Lagrangian function $f_{\theta_i}(\mathbf{\Phi}_s)$ regarding θ_i.
4. Iteratively perform steps (2)-(3) until the $\{\mathbf{\Phi}_s^*\}^T \mathbf{\Phi}_s^* = \boldsymbol{I}_P$ constraint is satisfied.

Step 2: Optimize common harmonic waves Ψ while fixing all individual harmonic waves $\{\Phi_s\}$: Given all adjusted individual harmonic waves $\{\mathbf{\Phi}_s\}$, estimating the common harmonic waves $\mathbf{\Psi}$ falls into solving the geometric median on the Stiefel manifold. The objective function of solving common harmonic waves $\mathbf{\Psi}$ becomes:

$$\begin{aligned} \min_{\mathbf{\Psi}} \boldsymbol{F}_{\mathbf{\Psi}} &= \min_{\mathbf{\Psi}} \sum_{s=1}^{m} d\left(\mathbf{\Phi}_s, \mathbf{\Psi}\right) \\ &= \min_{\mathbf{\Psi}} \sum_{s=1}^{m} \sqrt{P - Tr\left(\mathbf{\Phi}_s^T \mathbf{\Psi}\right)} \end{aligned} \tag{8}$$

Therefore, we can adopt the Weiszfeld algorithm [43] to efficiently solve the problem in (8) by alternately performing the following two steps until convergence:

1. Given the k-th estimated manifold center $\mathbf{\Psi}^{(k)}$ (purple circle in Fig. 2), the gradient of the energy function $d(\mathbf{\Phi}_s, \mathbf{\Psi}^{(k)})$ at $\mathbf{\Psi}^{(k)}$ on the Stiefel manifold can be calculated as $\nabla_{\mathbf{\Psi}^{(k)}} d(\mathbf{\Phi}_s, \mathbf{\Psi}^{(k)}) \in \mathcal{T}_{\mathbf{\Psi}^{(k)}}\mathcal{M}$, where $\mathcal{T}_{\mathbf{\Psi}^{(k)}}\mathcal{M}$ is the tangent plane at $\mathbf{\Psi}^{(k)}$ (shown as the green flat plane in Fig. 2). Therefore, we can obtain the manifold gradient $\nabla_{\mathbf{\Psi}^{(k)}} d(\mathbf{\Phi}_s, \mathbf{\Psi}^{(k)}) = (\mathbf{\Psi}^{(k)} \mathbf{\Phi}_s^T \mathbf{\Psi}^{(k)} - \mathbf{\Phi}_s) / \sqrt{P - Tr(\mathbf{\Phi}_s^T \mathbf{\Psi}^{(k)})}$ by [43] (blue solid arrow in Fig. 2). Finally, the mean tangent $\Delta\mathbf{\Psi}^{(k+1)} \in \mathcal{T}_{\mathbf{\Psi}^{(k)}}\mathcal{M}$ (red triangle in Fig. 2) is:

$$\begin{aligned} \Delta\mathbf{\Psi}^{(k+1)} &= -\nabla_{\mathbf{\Psi}^{(k)}} \sum_{s=1}^{m} d\left(\mathbf{\Phi}_s, \mathbf{\Psi}^{(k)}\right) \\ &= -\sum_{s=1}^{m} \frac{\mathbf{\Psi}^{(k)} \mathbf{\Phi}_s^T \mathbf{\Psi}^{(k)} - \mathbf{\Phi}_s}{\sqrt{P - Tr\left(\mathbf{\Phi}_s^T \mathbf{\Psi}^{(k)}\right)}} \end{aligned} \tag{9}$$

2. Map $\Delta\mathbf{\Psi}^{(k+1)}$ back to the Stiefel manifold by the Riemannian exponential map in (10) to obtain the updated manifold center (red circle in Fig. 2).

$$\mathbf{\Psi}^{(k+1)} = \exp_{\mathbf{\Psi}^{(k)}}\left(\gamma \Delta\mathbf{\Psi}^{(k+1)}\right) \tag{10}$$

It is worth noting that our proposed method penalizes the tangent vectors of outlier in the optimization process, as shown by the black solid arrow in Fig. 2, because they have a larger distance to the underlying manifold center. This property makes our proposed method more

robust to outlier influence than other approaches. Finally, the entire optimization scheme is summarized in the algorithm 1. We also discuss the parameter tuning, computational complexity analysis, and convergen analysis of our proposed method in section IV.

Algorithm 1: Parameters: $\beta = 0.1$; $\epsilon = 0.01$; $\varepsilon_1 = 0.01$; $\varepsilon_2 = 0.01$; $\gamma = 0.01$; $H = 100$.

Input: Adjacency matrix $\boldsymbol{W}_s \in \mathbb{R}^{N \times N}, s = 1, 2, \ldots, m$

Calculate Laplacian matrix $\boldsymbol{L}_s = \boldsymbol{D}_s - \boldsymbol{W}_s$

Initialize orthogonal matrix $\boldsymbol{\Phi}_s \in \mathbb{R}^{N \times P}$ through the Eigen-decomposition of Laplacian matrix $\boldsymbol{L}_s$

Initialize common network harmonic waves:

$\boldsymbol{\Psi} = eig(\frac{1}{m}\sum_{s=1}^{m} \boldsymbol{L}_s) \in \mathbb{R}^{N \times P}$

while ϵ is less than a pre-defined threshold ε_1 **do**
  for $s = 1 : m$ **do**
    for $i = 1 : H$ **do**
      Initialize parameter:
      $t = 1, LB_t = -1, \boldsymbol{\Phi}_s^t = \boldsymbol{\Phi}_s,$
      $\theta_i = \theta_{\min} + (i-1)\frac{\theta_{\max} - \theta_{\min}}{H}$
      while $\frac{f_\theta(\boldsymbol{\Phi}_s^t) - LB_t}{f_\theta(\boldsymbol{\Phi}_s^t) + 1} > \epsilon \parallel t < 100$ **do**
        $\widetilde{\boldsymbol{\Phi}}_s^t = \boldsymbol{U}(-\delta_{ij})_{N \times P}\boldsymbol{V}^T$
        where the $\boldsymbol{U}$ and $\boldsymbol{V}$ is from the single value decomposition $\nabla f_\theta(\boldsymbol{\Phi}_s^t) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.
        Update the lower bound:
        $LB_t = Tr(\nabla f(\boldsymbol{\Phi}_s^t)^T(\widetilde{\boldsymbol{\Phi}}_s^t - \boldsymbol{\Phi}_s^t)) + f(\boldsymbol{\Phi}_s^t)$
        Line search:
        $\alpha^* = \arg\min_{\alpha \in [0,1]} f_\theta(\boldsymbol{\Phi}_s^t + \alpha(\widetilde{\boldsymbol{\Phi}}_s^t - \boldsymbol{\Phi}_s^t))$
        $\boldsymbol{\Phi}_s^{t+1} = \boldsymbol{\Phi}_s^t + \alpha^*(\widetilde{\boldsymbol{\Phi}}_s^t - \boldsymbol{\Phi}_s^t), t = t + 1$
      end while
      if $\boldsymbol{\Phi}_s^{t\,T}\boldsymbol{\Phi}_s^t = I_P$ **then**
        $\boldsymbol{\Phi}_s^* = \boldsymbol{\Phi}_s^t$, break.
      end if
    end for
  end for
  Set start point $\boldsymbol{\Psi}^{(1)} = \boldsymbol{\Phi}_1^*$
  while $\Delta\boldsymbol{\Psi}^{(k)} < \varepsilon_2$ **do**
    $\Delta\boldsymbol{\Psi}^{(k+1)} = -\gamma\sum_{s=1}^{m}\frac{(\boldsymbol{\Psi}^{(k)}\boldsymbol{\Phi}_s^{*\,T}\boldsymbol{\Psi}^{(k)} - \boldsymbol{\Phi}_s^*)}{\sqrt{P - Tr(\boldsymbol{\Phi}_s^{*\,T}\boldsymbol{\Phi}^{(k)})}}$
    $\boldsymbol{\Psi}^{(k+1)} = \exp_{\boldsymbol{\Psi}^{(k)}}\Delta\boldsymbol{\Psi}^{(k+1)}, k = k + 1$
  end while
  Update $\boldsymbol{\Psi}^* = \boldsymbol{\Psi}^{(k+1)}$
  Compute $New_{cost} =$
  $\sum_{s=1}^{m}\{Tr(\boldsymbol{\Phi}_s^{*\,T}\boldsymbol{L}_s\boldsymbol{\Phi}_s^*) + \beta d^2(\boldsymbol{\Psi}^*, \boldsymbol{\Phi}_s^*)\}$
  $\epsilon = abs(New_{cost} - Old_{cost})$
  Update $Old_{cost} = New_{cost}$
end while

Output: Common harmonic waves $\boldsymbol{\Psi}^*$

C. Harmonic Waves Analysis

Empirical biomarkers (e.g., amyloid or tau deposition) are typically applied to predict early-stage AD in the traditional neuroimaging studies. Those empirical biomarkers can always be denoted as a column vector $f \in \mathbb{R}^N$ for each subject. Nowadays, manifold harmonic analysis is proposed to identify the harmonic-based alterations, which are related to the propagation pathways of neuropathological burdens. Therefore, the outlier-immunized common harmonic waves $\mathbf{\Psi}$ learned by our proposed method can be used to develop a new neuroimaging biomarker for each instance f by:

$$E(\psi_p) = |\langle f, \psi_p \rangle|^2 \tag{11}$$

Using the physics concept of harmonics, each $E(\psi_p)$ presents the harmonic energy propagating the neuropathology burden f through the underlying harmonic wave ψ_p. In Section III-C2, we demonstrate that our proposed new neuroimaging biomarkers achieve enhanced performance in group stratification than existing empirical biomarkers.

III. Experiments and Results

To evaluate the performance of our proposed harmonic analysis method, we compare the common harmonic waves optimized by our approach with the benchmark and the other groups of harmonic waves generated by the following methods: 1) Traditional method $\mathbf{\Psi}_o$: directly using the empirical biomarkers for classification; 2) Arithmetic mean $\mathbf{\Psi}_a$: simply averaging individual eigenvectors; 3) Pseudo manifold mean $\mathbf{\Psi}_p$: applying singular value decomposition (SVD) to the Laplacian matrix of the average adjacency matrix [30]; 4) Polynomial approximation $\mathbf{\Psi}_c$: the polynomial approximation of the average of Laplace-Beltrami operator [32]; 5) Fréchet mean $\mathbf{\Psi}_f$: estimating the Fréchet mean of all individual harmonic waves [35]. In addition, we refer to our method as the geometric median $\mathbf{\Psi}_g$.

In our experimental setting, we first measure the performance of our manifold learning method in estimating the common harmonic waves in terms of outliers in Sections III-B1 and III-B2. Then, the replicability of our learned common harmonic waves is evaluated in Section III-C1. Next, the diagnostic ability of our proposed new neuroimaging biomarkers in classifying CN, EMCI and LMCI subjects is investigated in Section III-C2. Furthermore, the outlier-immunized common harmonic waves are applied to discover harmonic-based alterations (significant neuroimaging biomarkers) in Section III-C3. Finally, we apply the general linear model (GLM) to evaluate the association between significant neuroimaging biomarkers and MMSE score in Section III-C4.

A. Image Acquisition and Data Preprocessing

We have collected neuroimaging data in the ADNI database, including T1-weighted magnetic resonance imaging (T1-weighted MRI) and diffusion tensor imaging (DTI) images from 94 subjects. First, according to a Desctrieux atlas [44], we parcellate the cortical surface into 148 cortical regions based T1-weighted MRI images. Second, a 148×148 anatomical connectivity matrix is produced by applying surface seed-based probabilistic

fiber tractography on the DTI images. Finally, the structural brain network is derived from the anatomical connectivity matrix, where the weight of the anatomical connectivity is defined by the number of fibers linking two brain regions normalized by the total number of fibers in the whole brain [44]. The outlier-immunized common harmonic waves $\mathbf{\Psi}$ are estimated by applying our proposed harmonic learning approach on above neuroimaging data. After obtaining the common harmonic waves $\mathbf{\Psi}$, there are another three ADNI datasets selected for group comparison. The imaging modalities of those three datasets are amyloid-PET, tau-PET and FDG-PET respectively, and the statistical information is in Table II. For PET imaging data, the cortical surface was parcellated into 148 structural brain regions, thus the standard update value ratio (SUVR) for each region is assembled into the column vector f.

B. Experiments on Synthetic Data

1) Evaluating the Accuracy of Common Harmonic Waves: Here, we are interested in evaluating the accuracy of identifying manifold centers vias different methods, including the arithmetic mean $\mathbf{\Psi}_a$, the Fréchet mean $\mathbf{\Psi}_f$, and our geometric median $\mathbf{\Psi}_g$. In this context, we generate a series of low-dimensional synthetic data by the following steps. We randomly synthesize fifteen three-dimensional orthonormal matrices as individual harmonic waves (in blue in Fig. 3) and two matrices as outliers (in yellow in Fig. 3), which can be represented as the points located on the Stiefel manifold (shown in Fig. 3(e)). Specifically, we first provide the ground truth (identity matrix), which is displayed in green in Fig. 3(a). Then, the orthonormal matrices are obtained by rotating the identify matrix with a given rotation axis and angles. The rotation angles are sampled from the Gaussian distribution with mean $\mu = 0$ and standard deviation $\xi = \pi/15$.

Since the adjacency matrices are not provided in synthetic data, the pseudo manifold mean $\mathbf{\Psi}_p$ is not appropriate for comparison here. Consequently, the manifold centers (common harmonic waves) evaluated via the arithmetic mean $\mathbf{\Psi}_a$, the Fréchet mean $\mathbf{\Psi}_f$, and our geometric median $\mathbf{\Psi}_g$ from above-mentioned seventeen orthonormal matrices (including two outliers) are exhibited in Fig. 3(b), (c) and (d), respectively. It is apparent that 1) the geometric median $\mathbf{\Psi}_g$ is much closer to the ground truth than the other manifold center estimation methods; 2) the arithmetic mean $\mathbf{\Psi}_a$ is outside the Stiefel manifold, indicating that the arithmetic mean is no longer an orthogonal matrix; and 3) the geometric median can quickly converge to the potential manifold center and be more robust to the outlier issue than the Fréchet mean, as illustrated in the purple (Fréchet mean) and red (geometric median) convergence trajectories in Fig. 3(e).

2) Evaluating the Robustness to Outliers: In this section, to further evaluate the performance of our method on the outlier issue, we construct an experiment to estimate the accuracy of the manifold center estimation by gradually increasing outliers. Specifically, we repeat the following procedure 100 times: 1) we generate 100 three-dimensional orthonormal matrices (including t% random outliers) based on the processing steps described in Section III-B1; 2) we apply our method and $\mathbf{\Psi}_f$ to these orthonormal matrices to estimate the manifold center; we calculate the accuracy by calculating the l_1-norm geodesic distance between the estimated center and ground truth. Finally, the accuracy (mean and

standard deviation) for estimating the manifold center with a proportion of outliers ranging from 0.02 to 0.35 is plotted in Fig. 4.

The experimental results demonstrate that 1) the l_1-norm geodesic distance between the manifold center detected by our geometric median and ground truth is significantly smaller ($p < 0.0001$) than that of the Fréchet mean, indicating that our method outperforms the Fréchet mean in identifying manifold centers; 2) with the increase in the outliers, the accuracy of the Fréchet mean drops sharply, while our method can maintain a high accuracy, indicating that our method has better robustness than the Fréchet mean in terms of outliers.

C. Experiments on the ADNI Dataset

1) Evaluating the Replicability of Common Harmonic Waves: In this experiment, we construct a reproducible experiment to evaluate the stability of our proposed method in discovering common harmonic waves. Specifically, we generate 100 test/retest datasets by the following resample procedure: 1) we select 60 brain networks from the total 94 brain networks as a common group and 20 subjects from the remaining 34 subjects for subsequent analysis; 2) we divide the 20 subjects into two specific groups, where each group includes 10 subjects; 3) we combine the common group and two specific groups to form two paired cohorts, each with 70 subjects. Finally, we evaluate the common harmonic waves by the proposed method on these two datasets independently. Since two cohorts have more than 80%(10/70) overlapped brain networks, we can assume that similar common harmonic waves should be obtained by applying our method to these two cohorts. In this context, we evaluate replicability by testing whether each element in the harmonic matrix was significantly different ($p < 0.01$) via the paired t-test. Since each specific brain region corresponds to a row in the harmonic matrix, we count the number of elements that fail the replicability test at each row and map them to the cortical surface, as shown in Fig. 5.

From the experimental results, we observe that 1) the pseudo manifold mean $\mathbf{\Psi}_p$ achieves the worst replicability performance; 2) geometric median $\mathbf{\Psi}_g$ has a smaller number of brain regions that fail the replicability test than the Fréchet mean $\mathbf{\Psi}_f$, indicating that geometric median has enhanced performance in replicability; 3) the arithmetic mean $\mathbf{\Psi}_a$ achieves similar replicability performance as our method, however, the orthogonality of arithmetic mean is often not guaranteed, as demonstrated in Fig. 3(b). Such nonorthogonality of the arithmetic mean will destroy the topological structure of the brain network, thereby affecting the classification performance, which will be demonstrated in Section III-C2; 4) though the polynomial approximation $\mathbf{\Psi}_c$ shows equivalent replicability as our method, the replicability of the left brain part of the polynomial approximation $\mathbf{\Psi}_c$ is far from the right brain part. Compared with the pseudo manifold mean $\mathbf{\Psi}_p$, the polynomial approximation $\mathbf{\Psi}_c$ greatly improves the replicability, which indicates that the polynomial approximation has a promoting effect on replicability.

2) Evaluating the Diagnostic Performance of Common Harmonic Waves: In this section, we estimate the diagnostic potential of our proposed neuroimaging biomarkers in CN, EMCI, and LMCI stratification. Since the traditional neuroimaging analysis methods widely use the region-wise SUVR score calculated from PET imaging data to predict

the early-stage AD, these empirical biomarkers can serve as benchmark to measure our neuroimaging biomarkers. In addition, the harmonic features extracted by the arithmetic mean $\mathbf{\Psi}_a$, the pseudo manifold mean $\mathbf{\Psi}_p$, the Fréchet mean $\mathbf{\Psi}_f$, and the polynomial approximation $\mathbf{\Psi}_c$ are also compared with our proposed method in this experiment. Specifically, the linear support vector machine (SVM) classifier is trained independently using the empirical biomarkers and neuroimaging biomarkers as input. Then, we can obtain the area under the receiver operating characteristic (AUROC) curve and area under the precision-recall (AUPR) curve scores using 10-fold cross-validation.

The Fig. 6 shows the classification results of different neuroimaging biomarkers and empirical biomarkers on three group comparisons (CN/EMCI, EMCI/LMCI, and CN/LMCI) in PET imaging data. It is clear that 1) the neuroimaging biomarkers based on geometric median $\mathbf{\Psi}_f$ achieve the best classification performance (highest AUROC and AUPR scores) over all other methods for three different PET imaging data, where star '*' or '**' indicates that our results are significantly better than those of the compared methods with $p < 0.1$ or $p < 0.01$ based on the between-area correlation; 2) the Stiefel means, including the Fréchet mean $\mathbf{\Psi}_f$ and the geometric median $\mathbf{\Psi}_g$, outperform other means estimated in Euclidean space; 3) the performance of polynomial approximation $\mathbf{\Psi}_c$ is similar as the Fréchet mean $\mathbf{\Psi}_f$, but it is still secondary to our approach; and 4) the AUROC and AUPR scores on FDG-PET data are lower than the other two PET imaging data (amyloid-PET and tau-PET). This is mainly because the FDG energy differences among CN, EMCI and LMCI are too small to influence the classification performance, which will be shown in Fig. 7. However, our proposed neuroimaging biomarkers still achieve higher classification accuracy than other methods and outperform the empirical biomarkers, indicating the great potential of applying our outlier-immunized common harmonic waves in the early diagnosis of AD.

3) Identifying Harmonic-Based Alterations on PET Imaging Data: Given the outlier-immunized common harmonic waves $\mathbf{\Psi}_g$ estimated by our proposed method, the new neuroimaging biomarkers can be extracted by (11). Since their effective classification ability has been validated in Section III-C2, we further investigate the harmonic-based alterations (significant neuroimaging biomarkers) that are related with the development and progression of AD.

For amyloid data, the total harmonic energy for each instance is firstly calculated via $E = \sum_{p=1}^{P} E(\psi_p)$, and then the total energy boxplot for the CN, EMCI and LMCI patient groups is shown in Fig. 7(a), where the total energy of the LMCI group is significantly higher ($p < 10^{-4}$) than that of the CN and EMCI groups. However, we cannot find a significant total energy difference ($p = 0.0369 > 0.01$) between the CN and EMCI patient groups. Moreover, we visualize the distribution of total harmonic energy in Fig. 7(b), with similar findings as the boxplot. Second, we measure the average energy of each harmonic wave for CN and LMCI group, as shown in the second and first layers in Fig. 7(c). Additionally, the corresponding Fisher score J_F is calculated to express the energy difference magnitude for each harmonic wave, which is illustrated in the third layer in Fig. 7(c), where the common harmonic waves showing the significant energy difference ($p < 0.01$) are tagged with star '★' Furthermore, the energy difference between CN and LMCI at

each harmonic wave is shown in Fig. 7(d). For the tau-PET imaging data, the total energy distribution and harmonic-specific energy difference are similar to the results of amyloid data, as shown in Fig. 7(e)-(h). These results demonstrate that the deposition of amyloid peptides and aggregation of tau tangles are associated with the progression of AD. In comparison, for the FDG-PET modality, we observe that the CN group has a significantly higher total energy ($p < 0.01$) than the LMCI group, while there is no significant total energy difference for the CN/EMCI and EMCI/LMCI comparisons (shown in Fig. 7(i)-(l)). These results show that the harmonic energy level of FDG can reflect the neurodegeneration of AD development. These significant harmonic waves (significant neuroimaging biomarkers) may play important roles in identifying the spreading of pathological burdens across brain networks.

4) Discovering Association Between Clinical Indicators and Significant Neuroimaging Biomarkers: As described in Section III-C3, some significant neuroimaging biomarkers based on our outlier-immunized common harmonic waves are detected via the *t*-test on PET imaging data, as shown in Fig. 7(c), (g) and (k). Therefore, we are interested in further research on whether there is a significant association between clinical indicators and our significant neuroimaging biomarkers In this experiment, theGLMis applied to predict the MMSE score using our neuroimaging biomarkers. We select the top 4 significant harmonic waves of each modality forGLManalyses and plot the statistical results in Fig. 8. It is clear that (1) the neuroimaging biomarkers based or the first common harmonic wave ψ_1 manifest a significant association ($p < 10^{-6}$) with the MMSE score on amyloid, tau and FDG data (shown in Fig. 8(a), (h) and (i)). Since the first common harmonic wave is a constant vector, its corresponding neuroimaging biomarker $E(\psi_1)$ represents the summation of the SUVR score. These results support the evidence that amyloid deposition, tau aggregation, and FDG levels are hallmarks of AD, which can be used for AD study; (2) The significan neuroimaging biomarkers of amyloid-PET (in the top of Fig. 8) anc tau-PET (in the middle of Fig. 8) are negatively correlated with MMSE scores, while the significant neuroimaging biomarkers of FDG-PET (in the bottom of Fig. 8) are positively associated with MMSE scores, which is consistent with the findings in Section III-C3. These results indicate the potential of the new neuroimaging biomarkers identified by our outlier-immunized common harmonic waves in predicting early AD cognitive decline.

IV. Discussion

A. Parameter Pruning

Our proposed method for estimating outlier-immunized common harmonic waves exists three parameters: the harmonic dimension parameter P, the hyperparameter β and the optimized parameter H.

As mentioned in Section II-A, since the higher frequency harmonic waves tend to be sensitive to potential noise, we determine to use only the smallest P eigenvectors $\mathbf{\Phi}_s^P$ instead of the whole eigenvectors $\mathbf{\Phi}_s$. According to the reconstruction loss calculated by matrix norm between original Laplacian matrix $\boldsymbol{L}_s$ and the reconstructed Laplacian matrix $\hat{\boldsymbol{L}}_s^P = (\mathbf{\Phi}_s^P)^T \mathbf{\Lambda}^P \mathbf{\Phi}_s^P$ with $\mathbf{\Lambda}^P$ as the diagonal matrix of the first P eigen-value of the Laplacian

matrix $\boldsymbol{L}_s$, the dimension P is determined, which is located at the stable point such that the reduction in reconstruction loss is marginal as P increases.

In terms of hyperparameter selection, the grid search method is applied to select the optimal β based on the classification accuracy of real data.

As we have discussed before, the main point of our algorithm is to find an appropriate θ for our solution, θ is more accurate if we have a bigger H, but when we select an oversize H, the time we cost will be too much, and the solution seems to be non-sensitive. Therefore, we determine the parameter by grid search method, we choose H from 10 to 150 with the interval of 10, and identify the optimal H based on the convergence performance of the algorithm.

Finally, we fix the harmonic dimension $P = 60$, the optimal parameter $\beta = 0.1$ and $H = 100$ in all our experiments.

B. Computational Complexity Analysis

The complexity of the Frank-Wolfe algorithm in algorithm 1 is highly related to that of SVD decomposition and matrix multiplication. In SVD decomposition, U and V are $(N \times N)$ and $(P \times P)$ matrices, respectively. Constructing these matrices has a time complexity of $O(N^3)$, while matrix multiplication has the same time complexity of $O(N^3)$. In the whole algorithm, the maximal iteration number is H, and the number of samples is m; therefore, optimizing algorithm 1 has the time complexity of $O(mHnN^3)$, where n is the iteration number of the Frank-Wolfe algorithm. Compared with the other methods, such as the interior-point solver SEDUMI, the Frank-Wolfe algorithm has the lowest time complexity [41], [42]. The inner loop of the algorithm 1 provides all updated individual harmonic waves $\{\boldsymbol{\Phi}_s\}$ for the external loop of the algorithm 1; thus, the whole algorithm 1 has a time complexity of $O(KmHnN^3)$ with K as the iteration number of the algorithm 1.

C. Convergence Analysis

As stated in Section II-B, the numerical scheme is divided into two steps. First, all the individual harmonic waves $\{\boldsymbol{\Phi}_s\}$ are solved by the Frank-Wolfe algorithm. By selecting a series of discrete θ, the objective function (7) is convergent according to Lemma 1 and Lemma 2 in Appendix A. For the set of different θ, we can obtain the convergent solution as proved in Lemma 3 in Appendix A. Second, we adopt the Weiszfeld algorithm to solve common harmonic waves $\boldsymbol{\Psi}$. Lemma 5 in Appendix A provides a detailed proof of the convergence of the objective function (8). Finally, we can prove that the whole algorithm is convergent, as shown in Theorem 6 in Appendix A.

D. Evaluation of Computational Efficiency and Accuracy

For the sake of studying the computational efficiency and accuracy of different neuroimaging biomarkers, we conduct an experiment examining the computational time and classification accuracy of those existing works with best performance values, arithmetic mean $\boldsymbol{\Phi}_a$, pseudo manifold mean $\boldsymbol{\Phi}_p$, polynomial approximation $\boldsymbol{\Phi}_c$, Fréchet mean $\boldsymbol{\Phi}_f$ and

geometric median $\mathbf{\Phi}_g$ on synthesized data. The synthesized data is generated by adding the gaussian distribution on the first-given two kinds of 148 × 1 vector, which can be regarded as two different categories. Notably, the outlier is the point that far away from those normal synthesized data. The results are shown in Table III. Those results show that 1) the computational efficiency of $\mathbf{\Phi}_a$ and $\mathbf{\Phi}_p$ have a significant superiority than the others; 2) the geometric median $\mathbf{\Phi}_g$ has achieved the best classification, the Fréchet mean $\mathbf{\Phi}_f$ take the second, and the geometric median $\mathbf{\Phi}_g$ has the least effect from the outliers, which indicates the robustness of our proposed approach to outliers.

E. Limitations and Future Work

Several limitations of the proposed method should be mentioned. (1) *Computational cost:* Although the proposed numerical scheme can theoretically guarantee convergence, it is computationally challenging to have large node numbers in the brain network analysis. In our future work, we plan to develop an effective optimization scheme to decrease the computational complexity while maintaining convergence. Such a method will be more reliable and reproducible for increasingly large brain networks. (2) *Small samples:* Compared to other methods, the experimental results demonstrate the enhanced performance of our estimated outlier-immunized common harmonic waves in the early AD diagnosis. However, our previously proposed common harmonic waves estimation method is still biased. This is mainly because estimating the common harmonic waves with a small number of brain networks in a high-dimensional manifold space is a challenging issue. Therefore, we will collect more brain networks to confirm the unbiased common harmonic waves estimation in the future. (3) *Global nature of harmonic waves:* Same as the Fourier bases, the common harmonic waves can be used as bases to project the brain signal to the frequency domain for the classification of AD. However, the global nature makes common harmonic waves impossible to discover disease-related brain regions and characterize the local neuropathological burdens propagation patterns. We have made some efforts in terms of harmonic localization [33], [45], however, these methods are still sensitive to outlier contamination. Thus, we will develop our proposed method into the localized common harmonic waves estimation version, which will be more powerful for brain network analysis.

V. Conclusion

In this paper, we propose a manifold learning method to estimate outlier-immunized common harmonic waves on the Stiefel manifold. Specifically, we measure the common harmonic waves as the geometric median of all individual harmonic waves. The outlier-immunized common harmonic waves offer a new window to discover the harmonic-based alterations related to AD propagation patterns and develop a new neuroimaging biomarker for forecasting early AD. We have demonstrated that our proposed method is more robust than other comparative approaches in dealing with outliers in synthetic data. In the extensive experiments on the ADNI dataset, our manifold learning approach achieves more consistent and reasonable results than existing methods that simply apply Euclidean operations on brain networks. In our future work, we plan to apply our outlier-immunized common harmonic waves to explore the potential pathogenic mechanism of other neurological disorders.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0112200, in part by the Key-Area Research and Development of Guangdong Province under Grants 2022A0505050014 and 2020B1111190001, in part by the Key-Area Research and Development Program of Guangzhou City under Grant 202206030009, in part by the National Natural Science Foundation of China under Grants U21A20520, 62172112, and 62102153, in part by the Natural Science Foundation of Guangdong Province of China under Grant 2022A1515011162, in part by the China Postdoctoral Science Foundation under Grant 2021M691062, and in part by Alzheimer's Disease Neuroimaging Initiative (ADNI).

Appendix A

Lemma 1: When $\theta \geq \theta_{\max} = \lambda_{\max} - \frac{\beta}{16\sqrt{2N\sqrt{P}}}$, $f_\theta(\mathbf{\Phi}_s)$ is a concave program, and only if $\theta \leq \theta_{\min} = \lambda_{\min} - \frac{\beta P}{8N\sqrt{(\nu)^3}}$, $f_\theta(\mathbf{\Phi}_s)$ is a convex conic program.

Proof: When $\theta \geq \theta_{\max}$, $\nabla f_\theta(\mathbf{\Phi}_s) \preceq 0$, $\nabla^2 f_\theta(\mathbf{\Phi}_s) \preceq 0$; when $\theta \leq \theta_{\min}$, $\nabla f_\theta(\mathbf{\Phi}_s) \succeq 0$, $\nabla^2 f_\theta(\mathbf{\Phi}_s) \succeq 0$. $\nabla f_\theta(\mathbf{\Phi}_s)$ and $\nabla^2 f_\theta(\mathbf{\Phi}_s)$ are shown as:

$$\begin{aligned} \nabla f_\theta(\mathbf{\Phi}_s) &= 2(\boldsymbol{L} - \theta \boldsymbol{I}_N)\mathbf{\Phi}_s - \frac{\beta \mathbf{\Psi}}{2\sqrt{P - Tr(\mathbf{\Phi}_s^T \mathbf{\Psi})}} \\ \nabla^2 f_\theta(\mathbf{\Phi}_s) &= 2(\boldsymbol{L} - \theta \boldsymbol{I}_N) - \frac{\beta P \boldsymbol{I}_N}{4n\sqrt{(P - Tr(\mathbf{\Phi}_s^T \mathbf{\Psi}))^3}} \end{aligned} \tag{12}$$

Consequently, $f_\theta(\mathbf{\Phi}_s)$ is a strictly concave function in terms of $\mathbf{\Phi}_s$ when $\theta \geq \theta_{\max}$, and $f_\theta(\mathbf{\Phi}_s)$ is a strictly convex function in terms of $\mathbf{\Phi}_s$ when $\theta \leq \theta_{\min}$.■

Lemma 2: The extreme point set of $\{\mathbf{\Phi}_s \in \mathbb{R}^{N \times P} \mid \mathbf{\Phi}_s^T \mathbf{\Phi}_s \preceq \boldsymbol{I}_P\}$ is

$$\{\mathbf{\Phi}_s \in \mathbb{R}^{N \times P} \mid \boldsymbol{I}_P \succeq \mathbf{\Phi}_s^T \mathbf{\Phi}_s, Tr(\mathbf{\Phi}_s^T \mathbf{\Phi}) = P\}$$

Proof: Let $\boldsymbol{X}_s$ be an extreme point of $\{\mathbf{\Phi}_s \in \mathbb{R}^{N \times P} \mid \mathbf{\Phi}_s^T \mathbf{\Phi}_s \preceq \boldsymbol{I}_P\}$. The singular value decomposition of $\boldsymbol{X}_s$ is $\boldsymbol{X}_s = \sum_{j=1}^{P} \sigma_j \boldsymbol{U}_j \boldsymbol{V}_j^T$, where $0 \leq \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_P \leq 1$. Thus, $\boldsymbol{X}_s^T \boldsymbol{X}_s = \sum_{j=1}^{P} \sigma_j^2 \boldsymbol{V}_j \boldsymbol{V}_j^T$ and $Tr(\boldsymbol{X}_s^T \boldsymbol{X}_s) = \sum_{j=1}^{P} \sigma_j^2$.

Supposing that $\sigma_1 < 1$, we define

$$\boldsymbol{Y} = (2\sigma_1 - 1)\boldsymbol{U}_1 \boldsymbol{V}_1^T + \sum_{j=2}^{P} \sigma_j \boldsymbol{U}_j \boldsymbol{V}_j^T, \boldsymbol{Z} = \boldsymbol{U}_1 \boldsymbol{V}_1^T + \sum_{j=2}^{P} \sigma_j \boldsymbol{U}_j \boldsymbol{V}_j^T$$

Then, we have $\boldsymbol{Y} \neq \boldsymbol{Z}, \boldsymbol{X} = \frac{1}{2}\boldsymbol{Y} + \frac{1}{2}\boldsymbol{Z}$, and

$$\boldsymbol{Y}^T\boldsymbol{Y} = (2\sigma_1 - 1)^2\boldsymbol{V}_1\boldsymbol{V}_1^T + \sum_{j=2}^{P}\sigma_j^2\boldsymbol{V}_j\boldsymbol{V}_j^T \preceq \sum_{j=2}^{P}\boldsymbol{V}_j\boldsymbol{V}_j^T = \boldsymbol{I}_P$$

$$\boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{V}_1\boldsymbol{V}_1^T + \sum_{j=2}^{P}\sigma_j^2\boldsymbol{V}_j\boldsymbol{V}_j^T \preceq \sum_{j=2}^{P}\boldsymbol{V}_j\boldsymbol{V}_j^T = \boldsymbol{I}_P$$

Therefore, $\boldsymbol{X}_s$ is not the extreme point, and the contradiction proves that $\sigma_1 = \sigma_2 = \cdots = \sigma_P = 1$, which completes the proof.■

Lemma 3: The algorithm in **Step 1** of Section II-B terminates in at most H iterations.

Proof: For the H-th iteration, $\theta_H = \theta_{\max}$; thus, $f_{\theta_H}(\boldsymbol{\Phi}_s)$ in the H-th iteration is a strictly concave function in terms of $\boldsymbol{\Phi}_s$.

Suppose $\boldsymbol{\Phi}_s^{1^T}\boldsymbol{\Phi}_s^1 \preceq \boldsymbol{I}_P$ and $\boldsymbol{\Phi}_s^{*^T}\boldsymbol{\Phi}_s^* = \boldsymbol{I}_P$; thus, $\boldsymbol{\Phi}_s^* - \boldsymbol{\Phi}_s^1 \neq 0$ and $f_{\theta_H}(\boldsymbol{\Phi}_s^1 + \alpha(\boldsymbol{\Phi}_s^* - \boldsymbol{\Phi}_s^1))$ is strictly concave with respect to the optimal step size α, estimated via line search strategy.

In addition, we have

$$\nabla f_{\theta_H}(\boldsymbol{\Phi}_s^1)^T\boldsymbol{\Phi}_s^* = \underset{\boldsymbol{\Phi}^T\boldsymbol{\Phi} \preceq \boldsymbol{I}_P}{\arg\min}\ Tr\left(\nabla f_{\theta_H}(\boldsymbol{\Phi}_s^1)^T\boldsymbol{\Phi}\right) \leq \nabla f_{\theta_H}(\boldsymbol{\Phi}_s^1)^T\boldsymbol{\Phi}_s^1$$

Thus, we conclude that

$$\frac{df_{\theta_H}(\boldsymbol{\Phi}_s^1 + \alpha(\boldsymbol{\Phi}_s^* - \boldsymbol{\Phi}_s^1))}{d\alpha}\Big|_{\alpha=0} = Tr\left(\nabla f_{\theta_H}(\boldsymbol{\Phi}_s^1)^T(\boldsymbol{\Phi}_s^* - \boldsymbol{\Phi}_s^1)\right) \leq 0$$

Therefore, $f_{\theta_H}(\boldsymbol{\Phi}_s^1 + \alpha(\boldsymbol{\Phi}_s^* - \boldsymbol{\Phi}_s^1))$ is strictly decreasing when $0 \leq \alpha \leq 1$. Consequently, when we take the optimal step size $\alpha^* = 1$, the optimal solution $\boldsymbol{\Phi}_s^t = \boldsymbol{\Phi}_s^{t-1} + \alpha^*(\boldsymbol{\Phi}_s^{t-1^*} - \boldsymbol{\Phi}_s^{t-1}) = \boldsymbol{\Phi}_s^{t-1^*}$ and hence $\boldsymbol{\Phi}_s^{t^T}\boldsymbol{\Phi}_s^t = \boldsymbol{I}_P$, which is the convergence criterion. Consequently, the algorithm in **Step 1** of Section II-B terminates in at most H iterations.■

Theorem 4: The algorithm in **Step 1** of Section II-B is convergent.

Proof: The dual question of solving the objective function (5) is $max_\theta\, min_{\boldsymbol{\Phi}_s^T\boldsymbol{\Phi} \preceq \boldsymbol{I}_P} f_\theta(\boldsymbol{\Phi}_s)$.

Specifically, for any $\theta_1 < \theta_2 < \cdots < \theta_i < \cdots \leq \theta_{\min}$,

$$
\begin{aligned}
f_{\theta_2}(\boldsymbol{\Phi}_s^{2*}) &= Tr\left(\boldsymbol{L}_s\boldsymbol{\Phi}_s^{2*}\boldsymbol{\Phi}_s^{2*^T}\right) + \beta\sqrt{Tr\left(\boldsymbol{I}_P - \boldsymbol{\Phi}_s^{2*^T}\boldsymbol{\Psi}\right)} \\
&\quad + \theta_2\left(P - Tr\left(\boldsymbol{\Phi}_s^{2*^T}\boldsymbol{\Phi}_s^{2*}\right)\right) \\
&> Tr\left(\boldsymbol{L}_s\boldsymbol{\Phi}_s^{2*}\boldsymbol{\Phi}_s^{2*^T}\right) + \beta\sqrt{Tr\left(\boldsymbol{I}_P - \boldsymbol{\Phi}_s^{2*^T}\boldsymbol{\Psi}\right)} \\
&\quad + \theta_1\left(P - Tr\left(\boldsymbol{\Phi}_s^{2*^T}\boldsymbol{\Phi}_s^{2*}\right)\right) \\
&\geq \min_{\boldsymbol{I}_P \succeq \boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s} Tr(\boldsymbol{L}_s\boldsymbol{\Phi}_s\boldsymbol{\Phi}_s^T) + \beta\sqrt{Tr(\boldsymbol{I}_P - \boldsymbol{\Phi}_s^T\boldsymbol{\Psi})} \\
&\quad + \theta_2\left(P - Tr\left(\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s\right)\right) = f_{\theta_2}(\boldsymbol{\Phi}_s^{2*})
\end{aligned}
$$

where $\boldsymbol{\Phi}_s^{i*}$ is the optimal solution of $min_{\boldsymbol{\Phi}_s^T\boldsymbol{\Phi} \preceq \boldsymbol{I}_P} f_{\theta_i}(\boldsymbol{\Phi}_s)$, so $f_{\theta_i}(\boldsymbol{\Phi}_s^{i*})$ is a strictly increasing function with respect to θ, and $f_{\theta_1}(\boldsymbol{\Phi}_s^{1*}) < f_{\theta_2}(\boldsymbol{\Phi}_s^{2*}) < \cdots < f_{\theta_i}(\boldsymbol{\Phi}_s^{i*})$. Therefore, the dual question has no optimal solution when $\theta < \theta_{\min}$.

In addition, $f_\theta(\boldsymbol{\Phi}_s)$ is a concave program when $\theta \geq \theta_{\max}$ in Lemma 1. Then, for $\theta_{\max} \leq \theta_1 < \theta_2 < \cdots < \theta_i$, the global minimizers of $\{f_{\theta_1}(\boldsymbol{\Phi}_s), f_{\theta_2}(\boldsymbol{\Phi}_s), \ldots, f_{\theta_i}(\boldsymbol{\Phi}_s)\}$ are all attained at an extreme point of the feasible region $\{\boldsymbol{\Phi}_s \in \mathbb{R}^{N\times P} \mid \boldsymbol{I}_P \succeq \boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s, Tr(\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}) = P\}$ in Lemma 2; thus, the optimal values of $\{f_{\theta_1}(\boldsymbol{\Phi}_s), f_{\theta_2}(\boldsymbol{\Phi}_s), \ldots, f_{\theta_n}(\boldsymbol{\Phi}_s)\}$ are all the same as the objective function (5). Therefore, when $\theta \geq \theta_{\max}$, there are infinitely many optimal solutions.

Therefore, these results suggest a continuous method to solve the dual question by obtaining the optimal solution of $max_\theta f_\theta(\boldsymbol{\Phi}_s^*)$ for the increasing sequence $\{\theta_i\} \in [\theta_{\min}, \theta_{\max}]$.

Combining Lemma 3, we know that the optimal solution of $max_\theta f_\theta(\boldsymbol{\Phi}_s^*)$ appears at most H iterations; therefore, the algorithm in **Step 1** of Section II-B is convergent.■

Lemma 5: The numerical scheme using the Weiszfeld updating is as follows:

$$
\begin{aligned}
\boldsymbol{\Psi}^{(k+1)} &= \exp_{\boldsymbol{\Psi}^{(k)}}\left(-\sum_{s=1}^{m}\nabla_{\boldsymbol{\Psi}^{(k)}}\boldsymbol{F}_{\boldsymbol{\Psi}}\right) \\
&= \exp_{\boldsymbol{\Psi}^{(k)}} = \left(-\sum_{s=1}^{m}\frac{(\boldsymbol{\Psi}^{(k)}\boldsymbol{\Phi}_s^T\boldsymbol{\Psi}^{(k)} - \boldsymbol{\Phi}_s)}{\sqrt{(P - Tr\,(\boldsymbol{\Phi}_s^T\boldsymbol{\Psi}^{(k)}))}}\right)
\end{aligned}
\tag{13}
$$

in solving:

$$
\begin{aligned}
\boldsymbol{F}_{\boldsymbol{\Psi}} &= \arg\min_{\boldsymbol{\Psi}}\sum_{s=1}^{m} d\,(\boldsymbol{\Phi}_s, \boldsymbol{\Psi}) \\
&= \arg\min_{\boldsymbol{\Psi}}\sum_{s=1}^{m}\sqrt{P - Tr\,(\boldsymbol{\Phi}_s^T\boldsymbol{\Psi})}
\end{aligned}
\tag{14}
$$

is convergent.

Proof: For the points $\boldsymbol{\Psi}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \ldots, \boldsymbol{\Phi}_s$ in the convex set of Stiefel manifold $\mathcal{M}$, we use $d(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$ to denote the geodesic distance between the two points $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ on the Stiefel manifold. The k-th iterations in the Weiszfeld algorithm:

$$F_{\Psi^{(k)}} = \sum_{s=1}^{m} d\left(\Psi^{(k)}, \Phi_s\right)$$

The corresponding function is defined on the tangent plane:

$$\widetilde{F}_{\widetilde{\Psi}^{(k)}} = \sum_{s=1}^{m} d\left(\widetilde{\Psi}^{(k)}, \widetilde{\Phi}_s\right)$$

$(\widetilde{\Psi}^{(k+1)}, \widetilde{\Phi}_s)$ is manifold point $(\Psi^{(k+1)}, \Phi_s)$ corresponding to the tangent plane point at $\Psi^{(k)}$. Note that $-\nabla_{\Psi} F_{\Psi} < 0$, and the exponential map is locally diffeomorphic onto a neighborhood of $\Psi^{(k)}$; therefore:

$$\widetilde{F}_{\widetilde{\Psi}^{(k+1)}} \leq \widetilde{F}_{\widetilde{\Psi}^{(k)}} = F_{\Psi^{(k)}} \tag{15}$$

The distance between two points on a positively curved manifold is smaller than the distance between their projections on the tangent plane according to Toponogov's theorem. This implies:

$$F_{\Psi^{(k+1)}} \leq \widetilde{F}_{\widetilde{\Psi}^{(k+1)}} \leq F_{\Psi^{(k)}}$$

The last inequality is from (15). because the objective function (14) is continuous, the objective function (14) is convergent.■

Theorem 6: The Algorithm 1 is convergent.

Proof: In Theorem 4, we show that **Step 1** in Section II-B is convergent, and the convergence of **Step 2** in Section II-B is proven in Lemma 5; therefore, it is easy to conclude that the Algorithm 1 is convergent.■

References

[1]. Mattson MP, "Pathways towards and away from Alzheimer's disease," Nature, vol. 430, no. 7000, pp. 631–639, 2004. [PubMed: 15295589]

[2]. Blennow K et al. , "Alzheimer's disease," Lancet, vol. 368, pp. 387–403, 2006. [PubMed: 16876668]

[3]. Patterson C et al., "World Alzheimer report 2018," Alzheimer's Dis. Int, London, U.K., pp. 6–7, 2018.

[4]. Brookmeyer R et al. , "Forecasting the global burden of Alzheimer's disease," Alzheimer's Dement., vol. 3, no. 3, pp. 186–191, 2007. [PubMed: 19595937]

[5]. Association A, "2015 Alzheimer's disease facts and figures," Alzheimer's Dement., vol. 11, no. 3, pp. 332–384, 2015. [PubMed: 25984581]

[6]. Hampel H and Lista S, "The rising global tide of cognitive impairment," Nature Rev. Neurol, vol. 12, no. 3, pp. 131–132, 2016. [PubMed: 26782338]

[7]. Petersen RC et al. , "Mild cognitive impairment: A concept in evolution," J. Intern. Med, vol. 275, no. 3, pp. 214–228, 2014. [PubMed: 24605806]

[8]. Lao H and Zhang X, “Regression and classification of Alzheimer’s disease diagnosis using NMF-TDNet features from 3D brain MR image,” IEEE J. Biomed. Health Inform, vol. 26, no. 3, pp. 1103–1115, Mar. 2022. [PubMed: 34543210]

[9]. Conturo TE et al. , “Tracking neuronal fiber pathways in the living human brain,” Proc. Nat. Acad. Sci, vol. 96, no. 18, pp. 10422–10427, 1999. [PubMed: 10468624]

[10]. Bihan DL, “Looking into the functional architecture of the brain with diffusion MRI,” Nature Rev. Neurosci, vol. 4, no. 6, pp. 469–480, 2003. [PubMed: 12778119]

[11]. He Y, chen Z, and Evans A, “Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer’s disease,” J. Neurosci, vol. 28, no. 18, pp. 4756–4766, 2008. [PubMed: 18448652]

[12]. Yao Z et al. , “Abnormal cortical networks in mild cognitive impairment and Alzheimer’s disease,” PLoS Comput. Biol, vol. 6, no. 11, 2010, Art. no. e1001006. [PubMed: 21124954]

[13]. Tijms BM et al. , “Single-subject grey matter graphs in Alzheimer’s disease,” PLoS One, vol. 8, no. 3, 2013, Art. no. e58921. [PubMed: 23536835]

[14]. Pereira JB et al. , “Disrupted network topology in patients with stable and progressive mild cognitive impairment and Alzheimer’s disease,” Cereb. Cortex, vol. 26, no. 8, pp. 3476–3493, 2016. [PubMed: 27178195]

[15]. Ling Q et al. , “A joint constrained CCA model for network-dependent brain subregion parcellation,” IEEE J. Biomed. Health Inform, vol. 26, no. 11, pp. 5641–5652, Nov. 2022. [PubMed: 35930507]

[16]. Stam CJ, “Modern network science of neurological disorders,” Nature Rev. Neurosci, vol. 15, no.10, pp. 683–695, 2014. [PubMed: 25186238]

[17]. Wu JW et al. , “Neuronal activity enhances tau propagation and tau pathology invivo,” Nature Neurosci., vol. 19, no. 8, pp. 1085–1092, 2016. [PubMed: 27322420]

[18]. Sepulcre J et al. , “Neurogenetic contributions to amyloid beta and tau spreading in the human cortex,” Nature Med., vol. 24, no. 12, pp. 1910–1918, 2018. [PubMed: 30374196]

[19]. Braak H and Del Tredici K, “Spreading of tau pathology in sporadic Alzheimer’s disease along cortico-cortical top-down connections,” Cereb. Cortex, vol. 28, no. 9, pp. 3372–3384, 2018. [PubMed: 29982389]

[20]. Vogel JW et al. , “Spread of pathological tau proteins through communicating neurons in human Alzheimer’s disease,” Nature Commun., vol. 11, no. 1, pp. 1–15, 2020. [PubMed: 31911652]

[21]. Rubinov M and Sporns O, “Complex network measures of brain connectivity: Uses and interpretations,” NeuroImage, vol. 52, no. 3, pp. 1059–1069, 2010. [PubMed: 19819337]

[22]. Brier MR et al. , “Network dysfunction in Alzheimer’s disease: Refining the disconnection hypothesis,” Brain Connectivity, vol. 4, no. 5, pp. 299–311, 2014. [PubMed: 24796856]

[23]. Aerts H et al. , “Brain networks under attack: Robustness properties and the impact of lesions,” Brain, vol. 139, no. 12, pp. 3063–3083, 2016. [PubMed: 27497487]

[24]. Naze S et al. , “Robustness of connectome harmonics to local gray matter and long-range white matter connectivity changes,” NeuroImage, vol. 224, 2021, Art. no. 117364. [PubMed: 32947015]

[25]. Sun Y et al. , “Inferring the individual psychopathologic deficits with structural connectivity in a longitudinal cohort of schizophrenia,” IEEE J. Biomed. Health Inform, vol. 26, no. 6, pp. 2536–2546, Jun. 2022. [PubMed: 34982705]

[26]. Vijayakumar N et al. , “Structural brain development: A review of methodological approaches and best practices,” Devlop. Cogn. Neurosci, vol. 33, pp. 129–148, 2018.

[27]. Raj A et al. , “Network diffusion model of progression predicts longitudinal patterns of atrophy and metabolism in Alzheimer’s disease,” Cell Rep., vol. 10, no. 3, pp. 359–369, 2015. [PubMed: 25600871]

[28]. Bassett DS and Sporns O, “Network neuroscience,” Nature Neurosci., vol. 20, no. 3, pp. 353–364, 2017. [PubMed: 28230844]

[29]. Raj A et al. , “A network diffusion model of disease progression in dementia,” Neuron, vol. 73, no. 6, pp. 1204–1215, 2012. [PubMed: 22445347]

[30]. Atasoy S et al. , “Human brain networks function in connectome-specific harmonic waves,” Nature Commun., vol. 7, no. 1, pp. 1–10, 2016.

[31]. Atasoy S et al. , “Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under LSD,” Sci. Rep, vol. 7, no. 1, pp. 1–18, 2017. [PubMed: 28127051]

[32]. Huang S-G, Lyu I, Qiu A, and Chung MK, “Fast polynomial approximation of heat kernel convolution on manifolds and its application to brain sulcal and gyral graph pattern analysis,” IEEE Trans. Med. Imag, vol. 39, no. 6, pp. 2201–2212, Jun. 2020.

[33]. Chen J et al. , “Characterizing the propagation pathway of neuropathological events of Alzheimer’s disease using harmonic wavelet analysis,” Med. Image Anal, vol. 79, 2022, Art. no. 102446. [PubMed: 35427899]

[34]. Fletcher PT et al. , “The geometric median on Riemannian manifolds with application to robust atlas estimation,” NeuroImage, vol. 45, no. 1, pp. S143–S152, 2009. [PubMed: 19056498]

[35]. Chen J et al. , “Learning common harmonic waves on Stiefel manifold–A new mathematical approach for brain network analyses,” IEEE Trans. Med. Imag, vol. 40, no. 1, pp. 419–430, Jan. 2021.

[36]. Chavel I, Eigenvalues in Riemannian Geometry. Cambridge, MA, USA: Academic Press, 1984.

[37]. Chikuse Y, Statistics on Special Manifolds, vol. 174. Berlin, Germany: Springer, 2012.

[38]. Ding Y et al. , “On equivalence of semidefinite relaxations for quadratic matrix programming,” Math. Operations Res, vol. 36, no. 1, pp. 88–104, 2011.

[39]. Ding Y and Wolkowicz H, “A low-dimensional semidefinite relaxation for the quadratic assignment problem,” Math. Operations Res, vol. 34, no. 4, pp. 1008–1022, 2009.

[40]. Xia Y and Han Y-W, “Partial Lagrangian relaxation for the unbalanced orthogonal Procrustes problem,” Math. Methods Operations Res, vol. 79, no. 2, pp. 225–237, 2014.

[41]. Ben-Tal A and Nemirovski A, “Non-euclidean restricted memory level method for large-scale convex optimization,” Math. Program, vol. 102, no. 3, pp. 407–456, 2005.

[42]. Nemirovski A, “Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddlepoint problems,” SIAM J. Optim, vol. 15, no. 1, pp. 229–251, 2004.

[43]. Aftab K, Hartley R, and Trumpf J, “Generalized weiszfeld algorithms for LQ optimization,” IEEE Trans. Pattern Anal Mach. Intell, vol. 37, no. 4, pp. 728–745, Apr. 2015. [PubMed: 26353290]

[44]. Destrieux C et al. , “Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature,” NeuroImage, vol. 53, no. 1, pp. 1–15, 2010. [PubMed: 20547229]

[45]. Chen J, Yang D, Cai H, Styner M, and Wu G, “Discovering spreading pathways of neuropathological events in Alzheimer’s disease using harmonic wavelets,” in Proc. 27th Int. Conf. Inf. Process. Med. Imag., 2021, pp. 228–240.

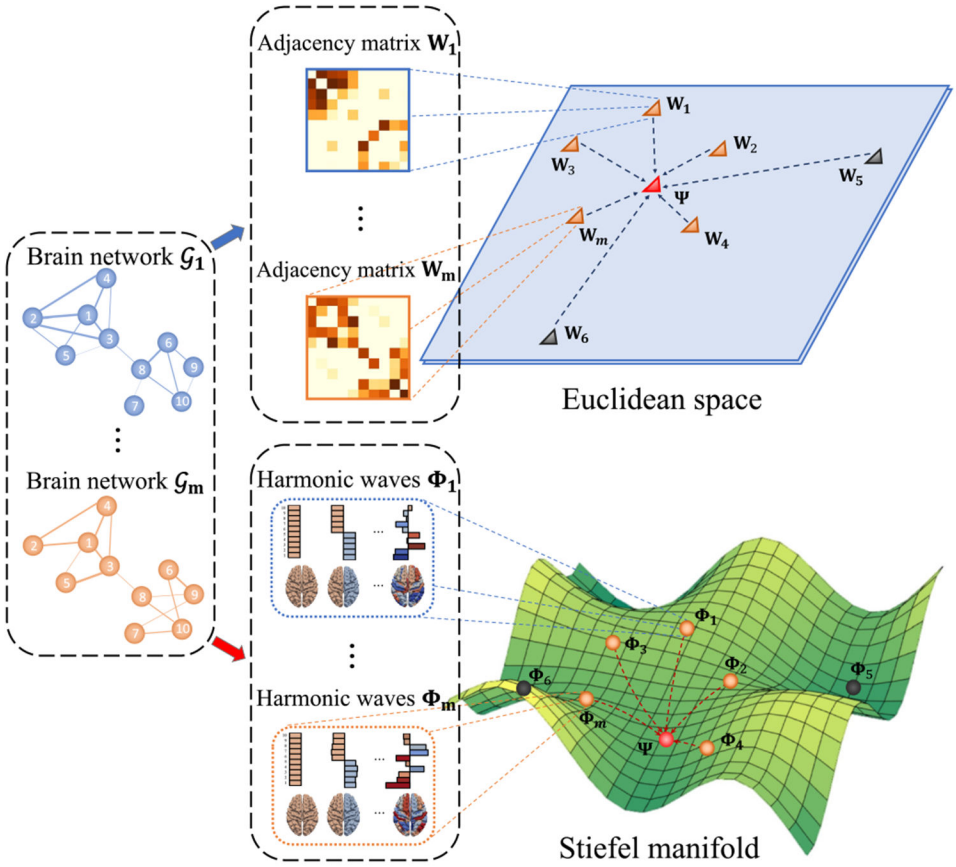

Fig. 1.

The traditional methods (blue arrow) identify the average brain network as a standard reference by calculating the arithmetic mean of all adjacency matrices on Euclidean space, which destroy the topological property of brain network. However, our manifold learning method (red arrow) estimates the outlier-immunized common harmonic waves on the Stiefel manifold by calculating the geometric median of all individual harmonic waves (yellow circle), which not only respects the brain network topology but is also insensitive to outliers (gray circle).

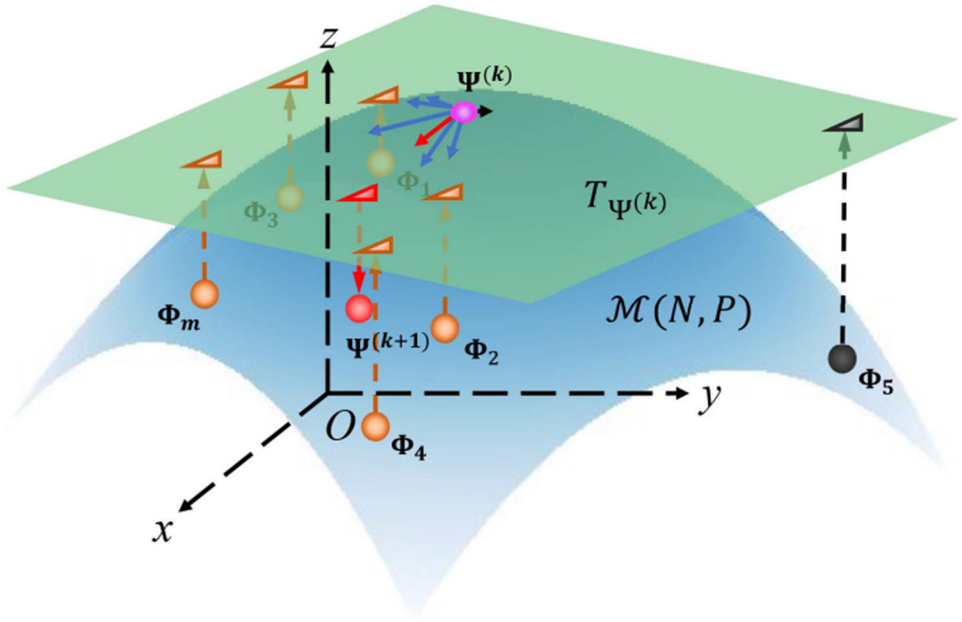

Fig. 2.

Illustration of optimizing common harmonic waves. Individual harmonic waves $\{\mathbf{\Phi}_s\}$ (yellow circle), an outlier (black circle), and k-th estimated common harmonic waves $\mathbf{\Psi}^{(k)}$ (purple circle) are resided on the Stiefel manifold $\mathcal{M}(N, P)$ (blue surface). The corresponding projected points (triangles) of individual harmonic waves $\{\mathbf{\Phi}_s\}$ are on the tangent plane $\mathcal{T}_{\mathbf{\Psi}^{(k)}}$ (green flat plane) of k^{th} manifold center $\mathbf{\Psi}^{(k)}$. The mean tangent (red triangle) is obtained by combining all gradient directions (solid arrow). Finally, map the mean tangent back to the updated manifold center $\mathbf{\Psi}^{(k+1)}$ (red circle) on the Stiefel manifold.

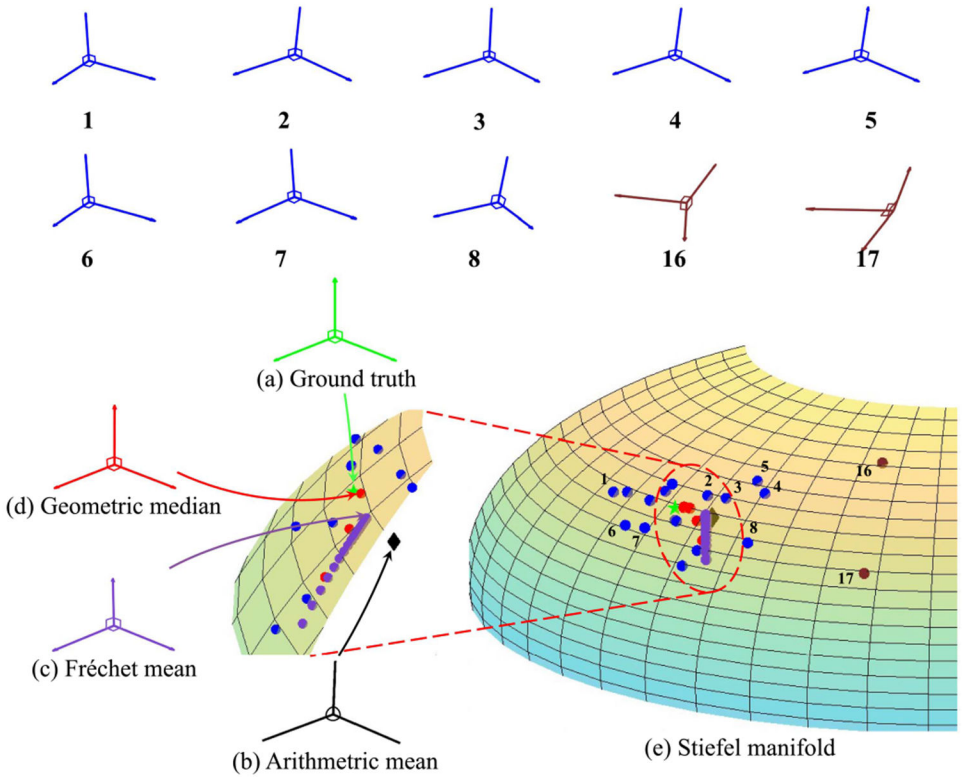

Fig. 3.

Illustration of manifold centers estimated via arithmetic mean $\mathbf{\Psi}_a$ (b), Fréchet mean $\mathbf{\Psi}_f$ (c) and geometric median $\mathbf{\Psi}_g$ (d). The top two rows show 8 orthonormal matrices (blue) constructed through rotating the identity matrix (a) with different rotation angles and 2 outliers (yellow). The evaluated results, all rotation matrices, and optimization process are displayed on the Stiefel manifold (e).

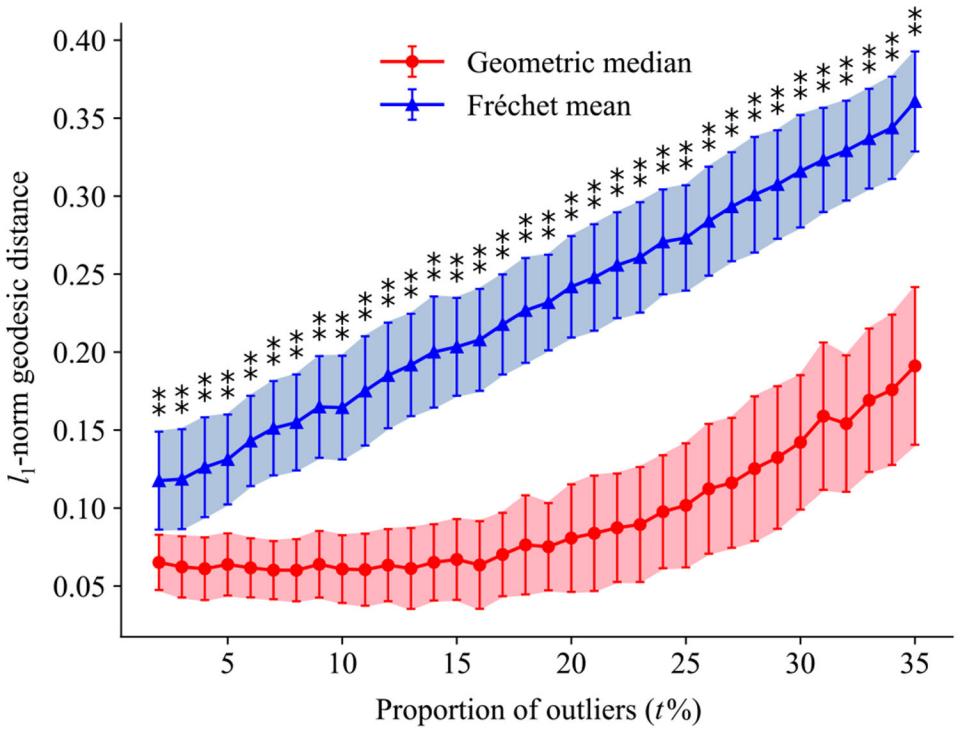

Fig. 4.

The robustness test results of geometric median (red curve) and Fréchet mean (blue curve) with regard to outliers. The black star '**' indicates that our results are significantly better ($p < 0.0001$) than the Fréchet mean.

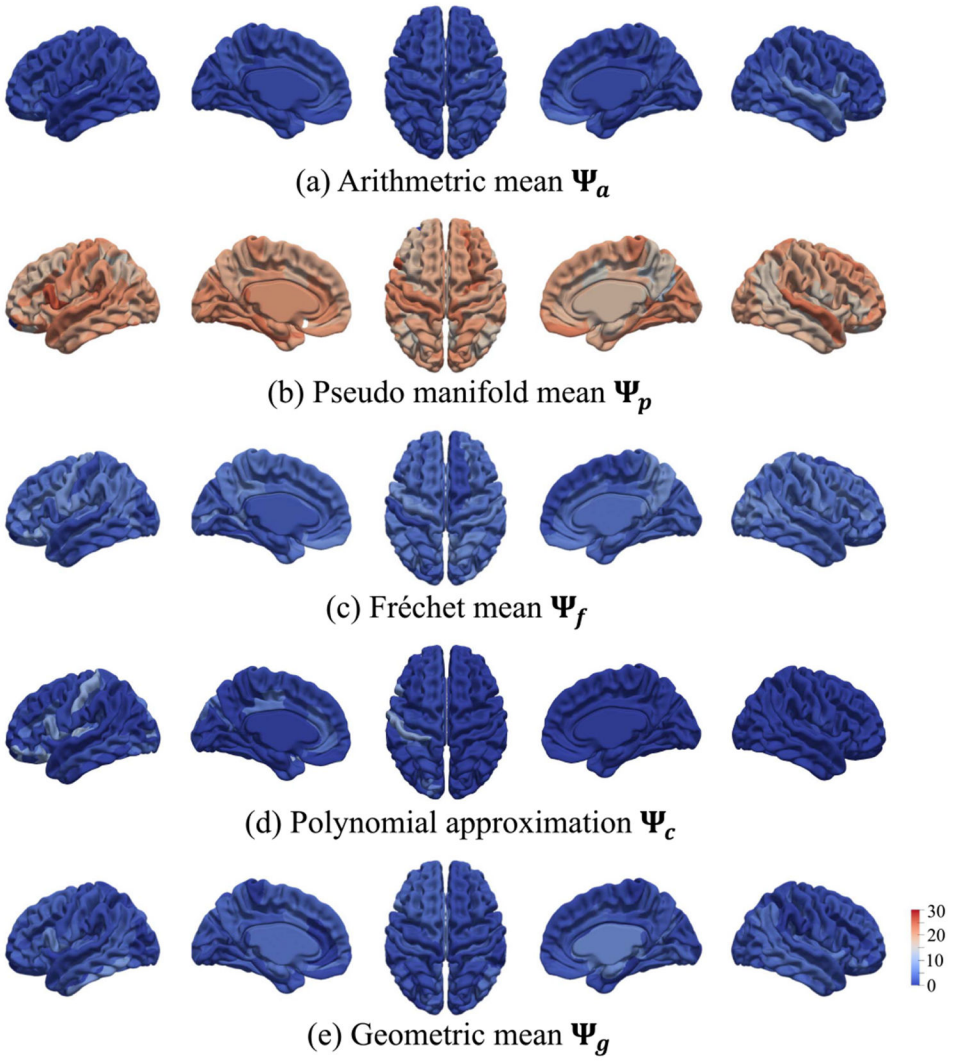

Fig. 5.

The replicability test results of (a) Arithmetic mean, (b) Pseudo manifold mean, (c) Fréchet mean, (d) Polynomial approximation and (e) Geometric median, where the number of failed replicability tests is reflected by the color of cortical surface.

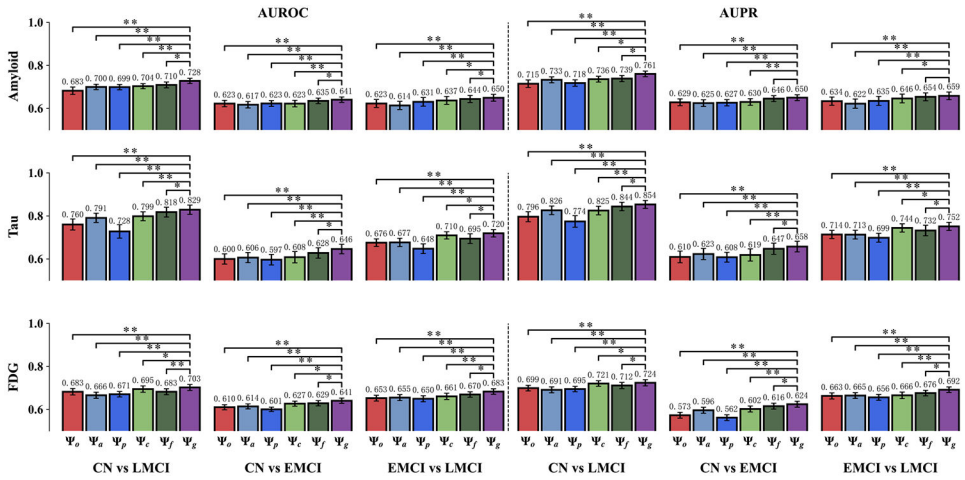

Fig. 6.

The classification results of using different neuroimaging biomarkers and empirical biomarkers on CN/EMCI, EMCI/LMCI, and CN/LMCI comparison in PET imaging data. The star '*' or '**' stands for the significant performance difference between our method and the other methods with $p < 0.1$ or $p < 0.01$ based on the between-area correlation.

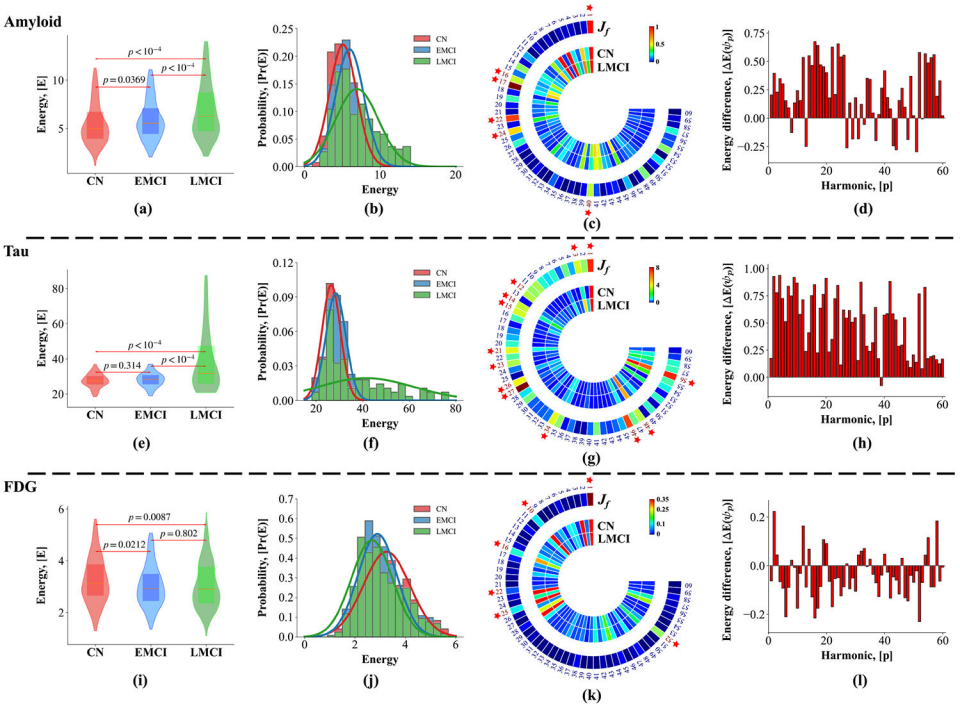

Fig. 7.

The harmonic-based alterations among CN, EMCI and LMCI discovered by our outlier-immunized common harmonic waves on amyloid-PET (top), tau-PET (middle) and FDG-PET (bottom) data. Top: (a)–(b) the total energy distribution of CN, EMCI, and LMCI on amyloid deposition. (c)–(d) the harmonic-specific energy difference between CN and LMCI group, where the significant harmonic waves are highlighted with star '★'. Middle: in the context of tau tangle. Bottom: in the context of FDG level.

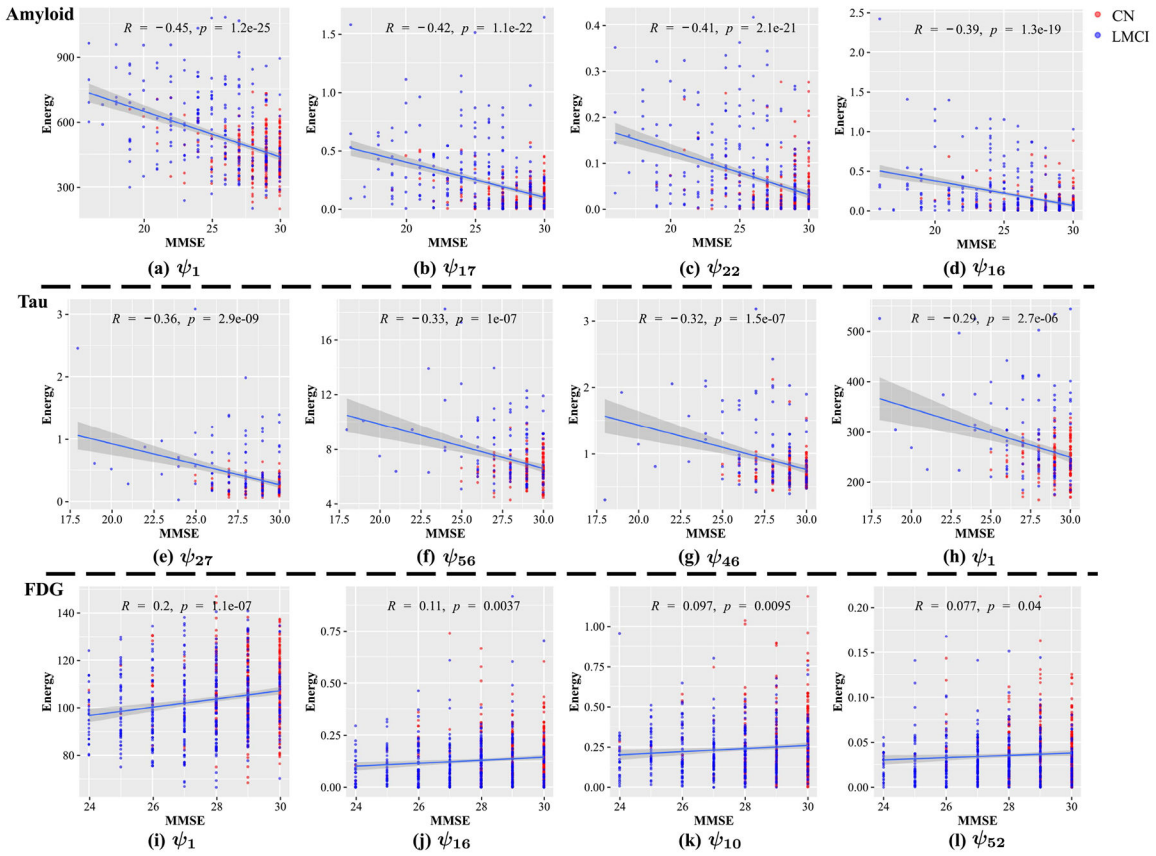

Fig. 8.

The association analyses between significant neuroimaging biomarkers and clinical indicator (MMSE) on amyloid-PET (top), tau-PET (middle) and FDG-PET (bottom) data. The subjects in CN and LMCI group are represented in red and blue dot, respectively. In addition, the statistical results (*R* value and *p* value) are also provided in each panel.

TABLE I

List of Notations

Notation	Remark
$\boldsymbol{L}_s$	The Laplacian matrix of the s-th graph $\mathcal{G}$
$\boldsymbol{\Phi}_s$	Individual harmonic waves of $\boldsymbol{L}_s$
$\boldsymbol{\Psi}$	Outlier-immunized common harmonic waves
$\boldsymbol{\Lambda}_s$	The augmented Lagrangian multiplier of $\boldsymbol{L}_s$
$\mathbb{R}^n$	n-dimensional real space
$\mathcal{M}$	The Stiefel manifold
$\mathcal{T}_{\boldsymbol{\Psi}}$, $\boldsymbol{\Delta}$	The tangent space and tangent vector on the Stiefel manifold at $\boldsymbol{\Psi}$
λ_{max}, λ_{min}	The maximum and minimum eigenvalues of Laplacian matrix $\boldsymbol{L}$
exp	The exponential map
$\nabla_{\boldsymbol{X}}\boldsymbol{F}$	The gradient of $\boldsymbol{F}$ at point $\boldsymbol{X}$ in the manifold space

TABLE II

Statistics of PET Imaging Data in Experiment

Number of Samples	Gender	Number	Range of Age	Average Age	CN	EMCI	LMCI
Amyloid	Male	450	55.0~91.4	73.4	136	184	130
	Female	389	55.0~89.6	71.7	148	145	96
	Total	**839**	**55.0~91.4**	**72.6**	**284**	**329**	**226**
Tau	Male	255	55.0~90.1	72.4	124	69	62
	Female	269	55.0~89.9	70.3	177	44	48
	Tatal	**524**	**55.0~90.1**	**71.3**	**301**	**113**	**110**
FDG	Male	592	55.0~91.4	73.9	169	182	241
	Female	472	55.0~89.6	72.2	166	148	158
	Total	**1064**	**55.0~91.4**	**73.1**	**335**	**330**	**399**

TABLE III

The Performance of Several Relevant Methods on Synthesized Data

Methods	Computational efficiency(s)	Accuracy	
		Without outliers	With outliers
Arithmetic mean Φ_a	**0.991**	0.884±0.013	0.643±0.026
Pseudo manifold mean Φ_p	1.476	0.891±0.012	0.647±0.027
Polynomial approximation Φ_c	20.083	0.921±0.013	0.689±0.026
Fréchet mean Φ_f	26.383	0.933±0.014	0.703±0.029
Geometric median Φ_g	28.760	**0.951±0.011**	**0.752±0.028**