

Coupled Harmonic Bases for Longitudinal Characterization of Brain Networks

Seong Jae Hwang¹ Nagesh Adluru⁴ Maxwell D. Collins¹ Sathya N. Ravi¹
Barbara B. Bendlin³ Sterling C. Johnson³ Vikas Singh^{2,1}

¹Dept. of Computer Sciences, University of Wisconsin – Madison

²Dept. of Biostatistics and Med. Informatics, University of Wisconsin – Madison

³William S. Middleton VA Hospital, Madison, WI ⁴Waisman Center, University of Wisconsin – Madison

Abstract

There is a great deal of interest in using large scale brain imaging studies to understand how brain connectivity evolves over time for an individual and how it varies over different levels/quantiles of cognitive function. To do so, one typically performs so-called tractography procedures on diffusion MR brain images and derives measures of brain connectivity expressed as graphs. The nodes correspond to distinct brain regions and the edges encode the strength of the connection. The scientific interest is in characterizing the evolution of these graphs over time or from healthy individuals to diseased. We pose this important question in terms of the Laplacian of the connectivity graphs derived from various longitudinal or disease time points — quantifying its progression is then expressed in terms of coupling the harmonic bases of a full set of Laplacians. We derive a coupled system of generalized eigenvalue problems (and corresponding numerical optimization schemes) whose solution helps characterize the full life cycle of brain connectivity evolution in a given dataset. Finally, we show a set of results on a diffusion MR imaging dataset of middle aged people at risk for Alzheimer’s disease (AD), who are cognitively healthy. In such asymptomatic adults, we find that a framework for characterizing brain connectivity evolution provides the ability to predict cognitive scores for individual subjects, and for estimating the progression of participant’s brain connectivity into the future.

1. Introduction

Large scale scientific initiatives such as the Human Connectome Project (HCP) are beginning to provide exquisite imaging data that may eventually enable a full structural connectivity mapping of the human brain [28]. For instance, diffusion magnetic resonance imaging (dMRI), an imaging modality central to the aforementioned studies, captures in a spatially localized (voxel-wise) manner, water diffusion

properties that can be used to infer the arrangement of network pathways in the brain [23, 27]. After suitable pre-processing, e.g., via so-called tractography procedures, we obtain a unique view of the fiber bundle layout that connects distinct brain regions [10] (see Fig. 1). From an analysis perspective, once the spatial organization of these fiber bundles is expressed as a graph whose nodes represent separate brain regions and the edges denote the “strength” of connection (e.g., number) of the connecting inter-region fibers, a variety of analyses can be conducted [12, 13, 24, 34]. For example, we may ask whether a specific *edge* of the graph exhibits statistically different connectivity measurements across clinically disparate groups: diseased and healthy. If the results of this hypothesis test are statistically significant, we can conclude that the corresponding fiber bundle (pertaining to the edge) is possibly affected by the disease.

But more recently, there is increasing interest in identifying not just imaging based or structural connectivity based biomarkers, rather to quantitatively characterize disease *progression* [3, 18, 31]. For example, studies may recruit subjects for multiple visits over a period of time (which varies from months to years) and acquire diffusion imaging data at several time points. In such *longitudinal* datasets, each subject (or sample) corresponds to multiple images (or the corresponding brain connectivities) at different time points. The scientific goal then is to identify the entire life cycle of brain connectivity evolution — from when a middle aged participant was healthy to a stage where the individual’s cognitive function has become much worse.

The standard approach to answering the question above is to characterize change in brain connectivity at the level of individual edges in the graph. There are two problems with this proposal. First, treating individual edges as primitives neglects the local context in which the edge exists in the actual object of interest — i.e., the entire connectivity graph of the individual. But from a technical perspective, a second issue is perhaps more important. For n regions, we obtain $\mathcal{O}(n^2)$ edges. After learning a model (e.g., associa-

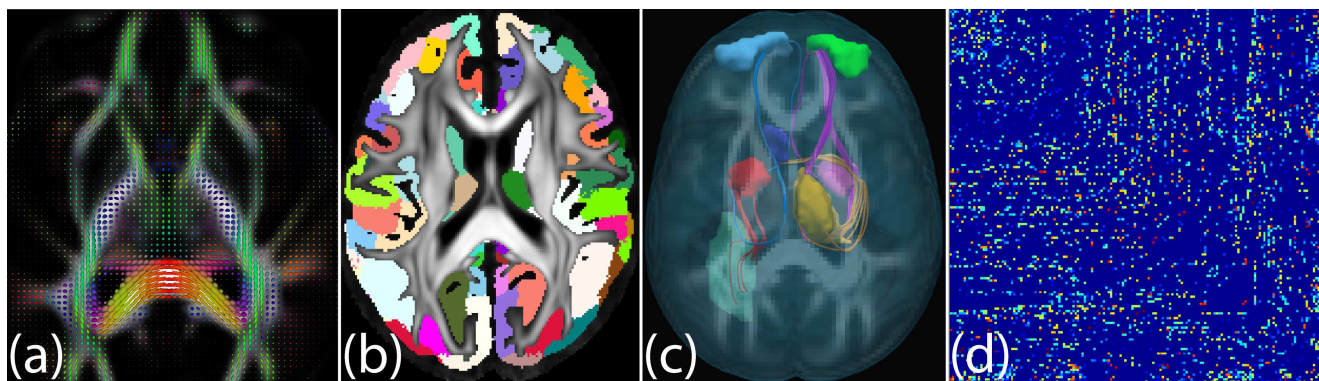


Figure 1: (a) Diffusion tensor ellipsoids obtained from the dMRI data using non-linear estimation. (b) Anatomical regions in cortical and sub-cortical gray matter are used to define the nodes in the brain network. (c) Fiber tracts (axonal pathways between brain regions) estimated via tractography are used to define connectivity strength between various gray matter nodes in the brain. (d) The brain networks can be represented as symmetric adjacency matrices.

tion of the connection weights with age) for each edge and estimating the significance of the fit (e.g., p -values), we *cannot* simply report the edges with small p -values as relevant. Since the analysis is being conducted for a large number of edges, the likelihood of false positives is high, so a severe multiple comparisons correction needs to be adopted. This correction may often be too conservative, and we may end up discarding connections that are, in fact, scientifically meaningful; this is an undesirable consequence of treating the edges individually. To avoid this problem, practitioners often rely on summary measures of the connectivity graph instead, such as clustering coefficients, small-worldness, modularity and so on [2, 20]. This works well but is limited in that we cannot uncover spatially localized effects of disease (or other covariates) on connectivity.

A more attractive solution is to think of the graphs as an object and utilize a suitable parameterization of the graph. One possibility is to use the Laplacians [19, 25], either at the level of individual subjects or in terms of distinct partitions of the full cohort progressively going from healthy to diseased, i.e., the first partition is comprised of completely healthy individuals whereas the k -th partition includes diseased subjects. Then, if we look at the full set of bases of the partition-specific Laplacian we can come up with ways to characterize change in these bases as we move from the healthy to the diseased partition. Of course, such a parameterization also enables *longitudinal* analysis. For example, if we have data for multiple time-points for each partition, we can track how the Laplacian bases evolve over time. To achieve these goals, i.e., for analyzing changes across partitions of disease severity or partition-specific longitudinal analysis, one requirement is the ability to derive a *coupled set of harmonic bases* for a set of ordered Laplacians (longitudinally and cross-sectionally). This allows treating the full data holistically *while preserving the ordinal nature of time and disease-induced cognitive decline*. While a mature body of literature in numerical analysis provides sophisticated ways of deriving orthonormal set of bases for

any self-adjoint operator, it provides little guidance on how to impose the coupling requirement, essential in this application. For example, we find that for most of the widely used eigenvalue decomposition methods [21, 32], it is non-trivial to modify the numerical scheme to satisfy the consistency requirement between consecutive set of eigen-bases. Addressing this limitation is a goal of this paper.

Key contributions. With the foregoing motivation, the core of this work deals with deriving efficient numerical optimization schemes to solve an ordinally and longitudinally coupled set of a generalized eigenvalue problems. To our knowledge, few publicly available alternatives currently exist [14, 15]. We provide (1) a novel formulation for estimating harmonic bases of brain connectivity networks that are smoothly varying in terms of both longitudinal as well as cross-sectional ordering (induced by a separate covariate such as cognitive performance). We provide an iterative numerical scheme for solving the problem using stochastic block coordinate descent based manifold optimization techniques. (2) We show that such a framework provides an exciting scientific tool in the following sense. Once the model has been estimated, we can vary a single parameter and “see” how the structural brain connectivity of an individual evolves over time or as a function of disease (see Fig. 2 for a qualitative demonstration). This yields a valuable mechanism for performing individual-level prediction. We show how our algorithm is able to provide connectivity prediction in a population of *healthy controls* who have some known risk factors of AD. Even though these individuals are *asymptomatic*, our approach is able to obtain a nominal degree of accuracy in assigning the subject to distinct cognitive quantiles. Demonstrating that we can, in fact, obtain a better than chance accuracy where the disease signal is so weak is the main contribution of this work.

2. Coupled harmonic bases for brain networks

Parameterization. Let us first describe a simple procedure for parameterizing the brain connectivity network in

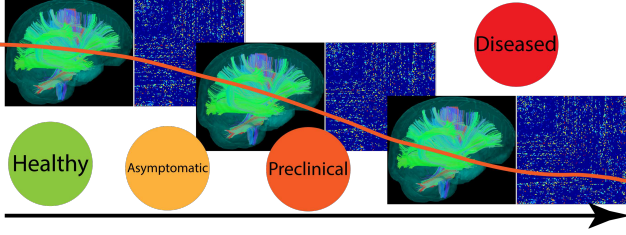


Figure 2: Using coupled harmonic representations derived from brain connectivity networks, we can turn a “knob” and see how a network evolves as the cognitive stage changes from healthy to diseased.

terms of its bases for individual subjects. Let A be a $n \times n$ weighted adjacency graph, A as in Fig. 1(d) representing a brain connectivity between n regions of the brain for a subject. We construct the Laplacian L , a commonly used tool/parameterization for representing graphs, defined as

$$L = D - A, \quad D(i, i) = \sum_{j=1}^n A(i, j),$$

where D is called the degree matrix. Based on spectral graph theory, the eigenvectors corresponding to lower order eigenvalues contain the ‘low frequency’ information which reflects the latent structure of the Laplacian. The bases are estimated by minimizing the following objective function

$$\min_{V \in \mathbb{R}^{n \times p}} \text{tr}(V^T L V), \quad \text{s.t. } V^T V = I, \quad (1)$$

where $\text{tr}(\cdot)$ is the trace functional. The solution V to the above numerical optimization problem consists of the eigenvectors associated with the p smallest eigenvalues of L which we solve to express a given brain connectivity network via its Laplacian and/or its p eigenvectors/eigenvalues.

Now suppose we are given a longitudinal dMRI dataset for N subjects with T time points: this provides NT brain networks. We can parameterize all the networks simultaneously by minimizing the following objective function

$$\min_{V_{[i,j]} \in \mathbb{R}^{n \times p}} \sum_{i=1}^N \sum_{j=1}^T \text{tr}(V_{[i,j]}^T L_{[i,j]} V_{[i,j]}) \quad (2)$$

$$\text{s.t. } V_{[i,j]}^T V_{[i,j]} = I,$$

where $L_{[i,j]}$ denotes the Laplacian matrix of brain network for subject i and time point j and $V_{[i,j]}$ denotes the set of p eigenvectors for $L_{[i,j]}$.

However, this formulation ignores a couple of key properties of our analysis goal (conceptually shown in Fig. 3). (1) Each subject has multiple time points which means that not all networks in the population are ‘independent’. There are strong dependencies among the networks derived from a single person observed over time. (2) The subjects can, if desired, be partitioned into distinct groups if a covariate of interest for the subjects is close enough (i.e., similar cognitive scores or a measure of pathology such as amyloid protein load may have roughly similar connectivity strength

[6]). This suggests that the bases that we find must also be related, or *coupled*, while still respecting, to the extent possible, their original Laplacians. If we consider the population of networks as a system, the recovery of the full set of bases must ensure a notion of consistency among $\{V_{[i,j]}\}$, governed by either the cognitive score grouping or longitudinal ordering described above.

We now present our proposed framework for adding constraints in (2) that will ensure that the full set of Laplacians are treated jointly. Without loss of generality, we can work with the example dataset scenario presented in Fig. 3.

2.1. Longitudinal coupling

In this section, we introduce basis coupling constraints that model the relationships (blue arrows in Fig. 3) between *temporally* consecutive bases. Suppose we consider the bases $V_{[\bullet,j]}$ and $V_{[\bullet,j+1]}$ for two consecutive time points j and $j+1$ for a specific subject. Since these are derived from the Laplacians of the same subject, we impose a homology constraint between the latent structures. In other words, we expect that $V_{[\bullet,j]}$ and $V_{[\bullet,j+1]}$ differ only by a small degree of *rotation*. Specifically, we impose

$$V_{[\bullet,j+1]} = R_{[j,j+1]}^\bullet V_{[\bullet,j]}, \quad (3)$$

where $R_{[j,j+1]}^\bullet \in \mathbb{SO}(n)$ which is a group of $n \times n$ orthogonal matrices with determinant = +1. This is, in fact, a Procrustes problem [30] of aligning the bases which provides the longitudinal evolution process of the set of bases as a sequence of rotation matrices. Note that the rotation matrix which aligns $V_{[\bullet,j+1]}$ to $V_{[\bullet,j]}$ is simply $R_{[j+1,j]}^\bullet = R_{[j,j+1]}^{\bullet T}$. Now, for eigenvectors $V_{[\bullet,j]}$ and $V_{[\bullet,j+1]}$, we see that

$$V_{[\bullet,j]}^T V_{[\bullet,j]} = V_{[\bullet,j]}^T R_{[j+1,j]}^\bullet V_{[\bullet,j+1]} = I, \quad (4)$$

$$V_{[\bullet,j+1]}^T V_{[\bullet,j+1]} = V_{[\bullet,j+1]}^T R_{[j,j+1]}^\bullet V_{[\bullet,j]} = I. \quad (5)$$

Multiplying the above two equations we have

$$\left(V_{[\bullet,j]}^T R_{[j+1,j]}^\bullet V_{[\bullet,j+1]} \right) \left(V_{[\bullet,j+1]}^T R_{[j,j+1]}^\bullet V_{[\bullet,j]} \right) = I$$

$$\implies V_{[\bullet,j]}^T \underbrace{\left(R_{[j+1,j]}^\bullet V_{[\bullet,j+1]} V_{[\bullet,j+1]}^T R_{[j,j+1]}^\bullet \right)}_{M'_{[j+1,j]}} V_{[\bullet,j]} = I.$$

Thus, we have now added a constraint on $V_{[\bullet,j]}$ that addresses the coupling between j and $j+1$ as

$$V_{[\bullet,j]}^T M'_{[j+1,j]} V_{[\bullet,j]} = I. \quad (6)$$

Note that for $j \notin \{1, T\}$, each $V_{[\bullet,j]}$ is tied to $V_{[\bullet,j-1]}$ and $V_{[\bullet,j+1]}$. Thus, the coupling matrices for those two relationships are $M'_{[j-1,j]}$ and $M'_{[j+1,j]}$ respectively. To account for both relations, we take the average coupling matrix as [7]:

$$M_{[\bullet,j]} = \frac{M'_{[j-1,j]} + M'_{[j+1,j]}}{2}. \quad (7)$$

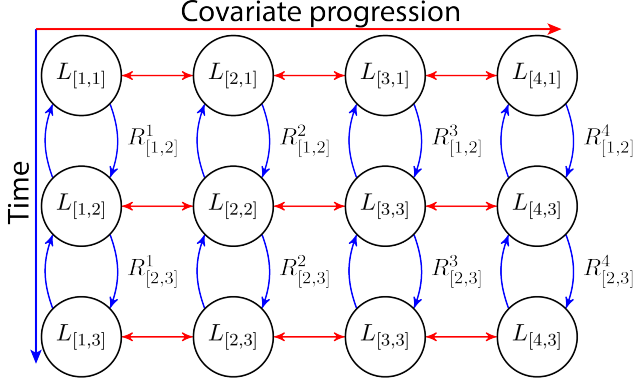


Figure 3: The graph representation of the coupled data matrices. The nodes in each row (cross-sectional) are coupled horizontally in red while the nodes in each column (longitudinal) are coupled vertically in blue.

For the boundary values of $j = 1$ and $j = T$, $M_{[\bullet,1]} = M'_{[2,1]}$ and $M_{[\bullet,T]} = M'_{[T,T-1]}$ respectively. Thus, for a subject i , we derive the following optimization model for discovering longitudinally coupled bases for the brain connectivity Laplacians,

$$\min_{V_{[i,j]} \in \mathbb{R}^{n \times p}} \text{tr}(V_{[i,j]}^T L_{[i,j]} V_{[i,j]}), \text{ s.t. } V_{[i,j]}^T M_{[i,j]} V_{[i,j]} = I. \quad (8)$$

Recall that the above constraint is the so-called generalized Stiefel constraint [1] with the *mass matrix* $M_{[i,j]}$. This is also known as the generalized eigenvalue problem that finds the first p smallest eigenvalues and the respective eigenvectors of $L_{[i,j]}$ for the mass matrix $M_{[i,j]}$. We can finally extend this implementation for all subjects as follows,

$$\min_{V_{[i,j]} \in \mathbb{R}^{n \times p}} \sum_{i=1}^N \sum_{j=1}^T \text{tr}(V_{[i,j]}^T L_{[i,j]} V_{[i,j]}) \quad (9)$$

$$\text{s.t. } V_{[i,j]}^T M_{[i,j]} V_{[i,j]} = I.$$

We point out that for each i and j , the above model is equivalent to (8).

2.2. Cross-sectional coupling

In this section, we present the appropriate constraints that will encode cross-sectional dependencies among the eigenvectors $\{V_{[i,j]}\}$. Let us say the population can be partitioned into K distinct groups (columns in Fig. 3) where each group/column is cognitively equivalent based on some battery of tests. Such partitions may also be derived by certain measures of pathology. For each group, we first construct an average Laplacian $X_{[i,\bullet]}$ at a fixed time point. This average Laplacian serves as a representative of that specific group. In principle, we can work with individual level Laplacians L , however, since the goal is to formulate the coupling of bases with respect to the covariates, the averaging helps us reduce the individual level variability and provides a more succinct picture of the network evolution along the trajectory of that covariate (e.g., cognitive scores).

Let us consider three such average Laplacians $X_{[i-1,\bullet]}$, $X_{[i,\bullet]}$ and $X_{[i+1,\bullet]}$ from three consecutive/ordered partitions. The corresponding eigenvectors will be $V_{[i-1,\bullet]}$, $V_{[i,\bullet]}$ and $V_{[i+1,\bullet]}$. Since these bases are derived from partitions with disjoint/distinct groups of subjects, we *cannot* assume a homological relationship between them. Since the coupling constraints will be added only between adjacent partitions it is nonetheless reasonable to assume that the bases will not change drastically affecting the full connectivity network. We encode this requirement as a sparsity constraint on the difference of the bases, e.g., via ℓ_0 norm. We use the relaxed ℓ_1 alternative,

$$g(V_{[i,\bullet]}) = \lambda (\|V_{[i-1,\bullet]} - V_{[i,\bullet]}\|_1 + \|V_{[i,\bullet]} - V_{[i+1,\bullet]}\|_1) \quad (10)$$

where $\lambda > 0$ is the regularization parameter which controls only the magnitude of the cross-sectional coupling.

Intuitively, it enforces similarities in certain dimensions of the cross-sectional bases while allowing the others to vary freely. In other words, we preserve some structural consistencies across the groups while still allowing the group-wise bases to be different. Note that for $i = 1$ and $i = K$, the regularization terms will only contain the first term and the second term of (10) respectively. The *cross-sectionally coupled* bases ($V_{[i,\bullet]}$) can then be estimated by minimizing the following,

$$\min_{V_{[i,\bullet]} \in \mathbb{R}^{n \times p}} \text{tr}(V_{[i,\bullet]}^T X_{[i,\bullet]} V_{[i,\bullet]}) + \lambda g(V_{[i,\bullet]}) \quad (11)$$

$$\text{s.t. } V_{[i,\bullet]}^T M_{[i,\bullet]} V_{[i,\bullet]} = I.$$

Putting it all together, we have the following *coupled* generalized eigenvalue formulation,

$$\min_{V_{[i,j]} \in \mathbb{R}^{n \times p}} \sum_{i=1}^K \sum_{j=1}^T \text{tr}(V_{[i,j]}^T X_{[i,j]} V_{[i,j]}) + \lambda \sum_{i=1}^{K-1} \sum_{j=1}^T \|V_{[i+1,j]} - V_{[i,j]}\|_1 \quad (12)$$

$$\text{s.t. } V_{[i,j]}^T M_{[i,j]} V_{[i,j]} = I; \quad V_{[i,j]} \in \mathbb{R}^{n \times p}.$$

We have now imposed both the longitudinal and cross-sectional basis coupling: for all partitions and time points.

3. Optimization scheme for coupled bases

In this section, we present an efficient numerical procedure for solving (12). Recall that the constraints involving the mass matrix form the generalized Stiefel manifold. We first present a few relevant technical details of the Stiefel manifold for completeness.

The set $\text{GF}_{n,p} = \{u \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p) : \text{rank}(u) = n\}$ of n -frames in \mathbb{R}^p is called the Stiefel manifold. When $p = n$, $\text{GF}_{n,n} := \text{GF}_n$ is the General Linear group or the set of $n \times n$ matrices with nonzero determinant. In other words, a Stiefel manifold is the set of $n \times p$ orthonormal matrices (with a Riemannian structure). $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ is the vector space of linear maps from \mathbb{R}^n to \mathbb{R}^p with the space

Algorithm 1 Stochastic block coordinate descent in $\text{GF}_{n,p}$

- 1: **Given:** $f : \text{GF}_{n,p} \rightarrow \mathbb{R}$, $V \in \text{GF}_{n,p}(M)$, $M \in \mathbb{R}^{n \times n}$
 - 2: **while** Convergence criteria not met **do**
 - 3: $\mathcal{S} :=$ Subproblem row indices
 - 4: $P_0 :=$ Initial feasible submatrix (19)
 - 5: $G :=$ Subdifferential of f w.r.t. P_0 (20)
 - 6: $W :=$ Descent curve in the direction of $-G$ on $\text{GF}_{s,p}(M_{\mathcal{S}\mathcal{S}})$ at P_0 (22)
 - 7: $\tau :=$ Step size under strong Wolfe conditions [17]
 - 8: $P :=$ Feasible point $W(\tau)$ of subproblem with sufficient decrease in f
 - 9: $V'(P) :=$ Update new feasible point (23)
 - 10: **end while**
-

$\mathbb{R}^{n \times p}$ of $n \times p$ matrices. Additional details on these topics such as tangent space, exponential map and retraction, are available in [1].

Our main strategy is to perform block coordinate descent over $\text{GF}_{n,p}$ to solve for $\{V_{[i,j]}\}$ for each matrix $X_{[i,j]}$ given in (12). Specifically, our algorithm iteratively decreases the objective value of the problem by finding the next feasible point in a curve which lies in the generalized Stiefel manifold $\text{GF}_{n,p}$ described in the constraints of (8), by adapting the scheme in [4, 9, 33]. This process is invoked as a module within our full pipeline.

Next, we show the entire framework of our algorithm which solves for all $V_{[i,j]}$ of the model (12) by iteratively solving for each $V_{[i,j]}$ while fixing the other decision variables $V_{[i',j']}$, $\forall i' \neq i \ \& \ j' \neq j$. In each iteration, we also update the mass matrix $M_{[i,j]}$ using the most recent bases.

Stochastic block coordinate descent in $\text{GF}_{n,p}$. For simplicity, let us focus on a single arbitrary partition and its Laplacian X , mass matrix M and the eigenvectors V and setup the following coupled model as in (11):

$$\begin{aligned} \min_{V \in \mathbb{R}^{n \times p}} \quad & \text{tr}(V^T X V) + \lambda g(V) \\ \text{s.t.} \quad & V^T M V = I, \end{aligned} \quad (13)$$

where the regularization term $g(V)$ from (10). First, we show how to solve (13) on a subset of the dimensions. This is a very common procedure in a coordinate descent method where the dimensions can be computed nearly independently to allow parallelization and make large-scale implementation possible. Specifically, we construct a subproblem for each submatrix $V_{\mathcal{S}} \in \mathbb{R}^{s \times p}$ where \mathcal{S} is a subset of s row indices of V . We choose this submatrix as the *free variable* which we ultimately solve for while fixing the complementary submatrix $V_{\bar{\mathcal{S}}}$ for the rows $\bar{\mathcal{S}}$ called the *fixed variable* which we essentially treat as constants. Assuming w.l.o.g. that $V_{\mathcal{S}}$ contains the first s rows of V and its complement $V_{\bar{\mathcal{S}}}$ contains the leftover indices, the con-

straint $V^T M V = I$ in (13) can be reorganized as

$$\begin{bmatrix} V_{\mathcal{S}} \\ V_{\bar{\mathcal{S}}} \end{bmatrix}^T \begin{bmatrix} M_{\mathcal{S}\mathcal{S}} & M_{\mathcal{S}\bar{\mathcal{S}}}^T \\ M_{\bar{\mathcal{S}}\mathcal{S}} & M_{\bar{\mathcal{S}}\bar{\mathcal{S}}} \end{bmatrix} \begin{bmatrix} V_{\mathcal{S}} \\ V_{\bar{\mathcal{S}}} \end{bmatrix} = I. \quad (14)$$

Rearranging it to move all the fixed variables on one side results in

$$V_{\mathcal{S}}^T M_{\mathcal{S}\mathcal{S}} V_{\mathcal{S}} + V_{\bar{\mathcal{S}}}^T M_{\bar{\mathcal{S}}\bar{\mathcal{S}}} V_{\bar{\mathcal{S}}} + V_{\bar{\mathcal{S}}}^T M_{\bar{\mathcal{S}}\mathcal{S}}^T V_{\mathcal{S}} =: \hat{H}, \quad (15)$$

for a constant matrix \hat{H} . With the full-rank assumption on $M_{\mathcal{S}\mathcal{S}}$, completing the square results the following:

$$\left(M_{\mathcal{S}\mathcal{S}}^{\frac{1}{2}} V_{\mathcal{S}} + M_{\mathcal{S}\mathcal{S}}^{-\frac{1}{2}} M_{\mathcal{S}\bar{\mathcal{S}}}^T V_{\bar{\mathcal{S}}} \right)^T \left(M_{\mathcal{S}\mathcal{S}}^{\frac{1}{2}} V_{\mathcal{S}} + M_{\mathcal{S}\mathcal{S}}^{-\frac{1}{2}} M_{\mathcal{S}\bar{\mathcal{S}}}^T V_{\bar{\mathcal{S}}} \right) = H \quad (16)$$

for a new constant matrix $H = \hat{H} + V_{\bar{\mathcal{S}}}^T M_{\bar{\mathcal{S}}\bar{\mathcal{S}}} M_{\mathcal{S}\mathcal{S}}^{-1} M_{\mathcal{S}\bar{\mathcal{S}}} V_{\bar{\mathcal{S}}}$. Since we assume that M is positive definite, $M_{\mathcal{S}\mathcal{S}}$ is also positive definite and invertible.

Now, given an orthogonal subproblem decision matrix P , the next feasible iterate can be provided as

$$V_{\mathcal{S}} = M_{\mathcal{S}\mathcal{S}}^{-\frac{1}{2}} P H^{\frac{1}{2}} - M_{\mathcal{S}\mathcal{S}}^{-1} M_{\mathcal{S}\bar{\mathcal{S}}}^T V_{\bar{\mathcal{S}}}, \quad (17)$$

which satisfies the constraints in (14). Note that by using a *retraction*, we can smoothly map the tangent vectors to the manifold and preserve the key properties of the exponential function necessary to perform feasible descent on the manifold. For the Stiefel manifold, a computationally efficient retraction comes from the Cayley transform which can be extended to the generalized Stiefel manifold shown by Equation (1.2) and Lemma 4.1 of [33]. Consequently, we can eliminate the extra computation of the matrix square roots and simplify (17) to the following:

$$V_{\mathcal{S}} = P - M_{\mathcal{S}\mathcal{S}}^{-1} M_{\mathcal{S}\bar{\mathcal{S}}}^T V_{\bar{\mathcal{S}}}. \quad (18)$$

If $P^T M_{\mathcal{S}\mathcal{S}} P = H$ for $M_{\mathcal{S}\mathcal{S}} \succ 0$ and non-singular H , the above equation satisfies the subproblem constraint. Thus, given the previous V , the descent curve starts at the point:

$$P_0 = V_{\mathcal{S}} + M_{\mathcal{S}\mathcal{S}}^{-\frac{1}{2}} M_{\mathcal{S}\bar{\mathcal{S}}}^T V_{\bar{\mathcal{S}}}. \quad (19)$$

So far, we have shown how to setup the initial point for the line search step. Next, we describe how to compute the descent curve of the subproblem on the generalized Stiefel manifold for the line search on the manifold. The first step is to find the gradient of the objective function of (13) which is $f(V) = V^T X V + \lambda g(V)$. Thus, for $f(U)$ and $V(P)$ where the next feasible point V as a function of P as in (18), the gradient of $f \circ V(P)$ w.r.t. P is

$$\frac{\partial (f \circ V(P))}{\partial P} = 2(X_{\mathcal{S}\mathcal{S}} V_{\mathcal{S}} + X_{\mathcal{S}\bar{\mathcal{S}}} V_{\bar{\mathcal{S}}}) + \lambda g'(V(P)), \quad (20)$$

where $g'(V(P))$ is the subgradient of the regularization term (10). Next, we project the descending subgradient $-G'$ at P_0 onto the tangent space of the manifold $\text{GF}_{i,p}(M_{\mathcal{S}\mathcal{S}})$ by constructing a skew-symmetric matrix:

$$Q = G P_0^T - P_0 G^T, \quad (21)$$

Algorithm 2 Coupled bases framework using SBCD

```
1: Given:  
    $f : \text{GF}_{n,p} \rightarrow \mathbb{R}, V_{[:,i]} \in \text{GF}_{n,p}(M_{[:,i]}), M_{[:,i]} \in \mathbb{R}^{n \times n}$   
2: while Convergence criteria not met do  
3:   for  $i = 1, \dots, K$  do  
4:     for  $j = 1, \dots, T$  do  
5:        $V_{[i,j]} :=$  Free variable  
6:        $V_{[i,j]} :=$  SBCD( $V_{[i,j]}$ ) (Alg. 1)  
7:     end for  
8:     for  $j = 1, \dots, T$  do  
9:        $R_{[i,j]} :=$  Rotation matrix (3)  
10:    end for  
11:    for  $j = 1, \dots, T$  do  
12:       $M_{[i,j]} :=$  Mass matrix (6), (7)  
13:    end for  
14:  end for  
15: end while
```

which conveniently allows the Cayley transform as in [33] to smoothly map from the tangent space to the generalized Stiefel manifold to create the descent curve W as a function of τ :

$$W(\tau) = \left(I + \frac{\tau}{2} Q M_{SS} \right)^{-1} \left(I - \frac{\tau}{2} Q M_{SS} \right) P_0. \quad (22)$$

We can linearly search over the descent curve to find the new point $P = W(\tau)$ for some τ which results sufficient decrease in f . Thus, the next feasible decision variable $V' \in \mathbb{R}^{n \times p}$ as a function of P is

$$V'(P) = \begin{bmatrix} P - M_{SS}^{-1} M_{SS}^T V_{\bar{S}} \\ V_{\bar{S}} \end{bmatrix} \quad (23)$$

so we can finally assign the current V with V' . By the construction of the coupled bases model (13), V' is updated to minimize the objective function *while it remains coupled* with the other longitudinal and cross-sectional bases connected to V . The pseudo-code of the algorithm is in Alg. 1.

Iterative SBCD in $\text{GF}_{n,p}$ for bases coupling. With the stochastic block gradient descent (SBCD) method roughly similar to [16] as a solver for a single coupled basis, we now setup the framework to solve for all coupled bases. Specifically, given multiple matrices $X_{[i,j]}$ for $i = 1, \dots, K$ and $j = 1, \dots, T$, we set up grid-like iterations as shown in Alg. 2 where we iteratively pass through all possible pairs of i and j . In each iteration, we set $V_{[i,j]}$ for the current pair of i and j to be the free variable and set $V_{[i',j']}$ of the remaining $i' \neq i$ and $j' \neq j$ to be the fixed variables. We solve for only the free variable $V_{[i,j]}$ using SBCD iteratively for all i and j .

Since (13) imposes the longitudinal coupling based on the mass matrices that are precomputed from the bases available at that iteration, they might not reflect the most accurate longitudinal relations precomputed mass matrices

involving rotation matrices of the bases. Therefore, we must update the mass matrices so that they reflect the most accurate bases (encoding the naturally derived longitudinal trajectory) by recomputing the rotation matrices of the newly updated bases. Thus, the mass matrix computation step immediately follows the bases update step. We repeat these steps iteratively until convergence criteria are met.

4. Experiments

Data and scientific goal. Our dataset focuses on a cohort of middle aged individuals who are at risk for Alzheimer’s disease (AD) due to a positive family history (at least one parent with confirmed diagnosis of AD). Our data corresponds to at least *three* longitudinal scans of these subjects. The participants are *cognitively healthy* but some AD related brain changes, while subtle, have already begun. Note that in the analysis of anatomical changes in standard magnetic resonance images (*not* brain connectivity as we do here), numerous findings suggest that accurate quantification of brain “changes”, e.g., via tensor based morphometry, is often more sensitive than the analysis of individual images independently [8]. But in the context of AD (and for other brain disorders) most, if not all, such studies focus on data that cover the *full spectrum of the disease* (healthy to AD) – the disease effects of those data are *much stronger* and arguably easier to detect than those in the setting we consider here. We expect that estimating the longitudinal change process in connectivity accurately via the coupled harmonic bases model will enable identifying a *disease signal* even in the pool of *healthy* (but at-risk) individuals.

Workflow for deriving brain connectivity networks.

There are three key steps in deriving brain connectivity networks for a given population study. **(a) Coordinate system.** For population level analysis of brain images, one typically needs to register all the images onto a standard coordinate system in a way that avoids any unwanted biases. We follow recommended procedures for deriving an unbiased coordinate system for the 3D+time regime as follows [11]. We first estimate a subject-specific average that is temporally unbiased. The subject specific averages are then used to generate an unbiased population level average template space, the process is summarized in Fig. 4. **(b) Edges.** We use tractography for deriving edges of the *in vivo* brain (structural) networks. The key ingredient needed for these algorithms is the orientational information of white matter fibers passing through a voxel, inferred by fitting a tissue model to the acquired MR signal from diffusion weighted imaging. Our data was acquired at a single shell diffusion weighting of $b = 1000$ s/mm². Although limited in its ability to resolve crossing fiber tissue, for this data the most reliable model that can be fit is the so called diffusion tensor. The principal eigenvector of the diffusion tensor in a voxel

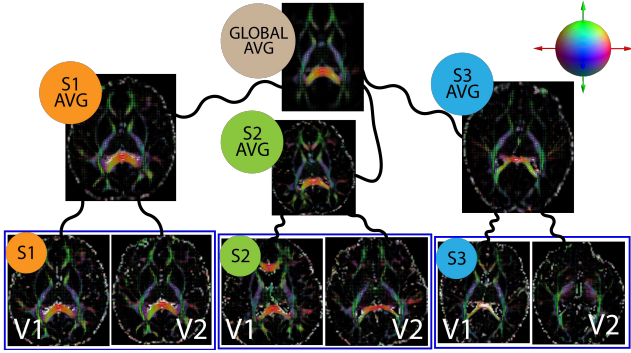


Figure 4: Unbiased estimation of the global coordinate system for the longitudinally acquired imaging data. Visits are averaged first which are then used to estimate the global average. Each of the curved black lines represents a combination of rigid, affine and non-linear diffeomorphic transformations. These transformations and the spatial averages are estimated iteratively. Diffusion tensors are directly registered using log-Euclidean framework [35].

provides a proxy to the predominant orientation of the white matter fibers in that voxel. Using this information, we repeated probabilistic tractography twenty times [5]. The normalized standard deviation of the eigenvectors of the tensor (also known as fractional anisotropy) in the tracts passing between two regions serves as the edge strength between nodes in our experiments. **(c) Nodes.** The third key component is the node definition. For defining nodes, we relied on expert neuroanatomy groups who have carefully delineated the boundaries of regions in the brain based on their knowledge (including histological studies). We used the gray matter atlas defined on a DTI template [29] which provides 160 distinct regions in the brain.

Covariates for creating sub-cohorts/partitions. We partition the full cohort using two key ordinal covariates: Rey Auditory Verbal Learning Test (RAVLT) [22] and Mini Mental State Exam (MMSE) [26]. These are cognitive performance scores based on tests that assess the cognitive functioning of the subjects and common in preclinical assessments of AD. Note that there is a systematic effect of age and gender on these scores. To control for these nuisance variables, we perform regression against these variables and derive z -scores for both RAVLT and MMSE. This imposes an implicit ordering of the subjects for the two measures, after the effect of age and sex has been accounted for. We derive K partitions of the z -scores to “stage” the cognitive status (and the subjects) into distinct cognitive quantiles. This implies that even within the full set of “healthy” individuals, subjects that fall within the same quantile are similar. If this staging is finer, we have fewer samples in each partition. We used $K \in \{2, 3, 4\}$ partitions to keep the individual-level variance manageable while still allowing us to identify the general connectivity evolution patterns. Although we assume that all participants

are recruited into the study concurrently (which may not be true), since they are assigned to distinct cognitive quantiles, estimating their longitudinal trajectories is reasonable (also, since the entire cohort is middle aged). In case of uneven distribution along the time axis, standard imputation strategies may be needed. Here, we only utilized data where all three time points were available to keep the presentation simple and avoid concerns related to the potential effect/bias of the specific imputation methods. Once the subjects are assigned to the appropriate partitions based on their z -scores (columns in Fig. 3), we can derive the average Laplacians, formulate the system as (12) and solve for the coupled bases.

Design. We use 68 subjects with three longitudinal time points and partition them for different K based on RAVLT and MMSE z -scores. Then, for each setting, we compute four sets of bases: **(a)** non-coupled (2), **(b)** longitudinally coupled (9), **(c)** cross-sectionally coupled ((12) with matrix $M = I$) and **(d)** longitudinally+cross-sectionally coupled (12). Thus, each partition now has a set of longitudinal bases (vertical direction in Fig. 3). For a novel test subject, we can calculate the corresponding connectivity graph and compare to each of the K partitions. The quantile of the closest partition is the label of this new subject. First, to measure the overall accuracy of this procedure, we use 21 ‘held out’ test subjects, which were *not* used to compute the four sets of bases to avoid overfitting, where each subject has three longitudinal scans available. So, we have total of $21 \times 3 = 63$ distinct Laplacians. We perform two classification tasks. First, we only predict the quantiles of the Laplacians at the *first* (or baseline) time point. Next, we predict all 63 Laplacians which is expected to be much harder since the quantiles of the subsequent time points are not used in deriving the partitions. Nonetheless, we expect that if our coupled bases are accurate, the information from the first time point should, in principle, affect subsequent time points in a way that allows our model to still predict the label correctly. We evaluated $p \in \{n/4, n/2, 3n/4\}$ and chose $p = n/4$ since the residual is extremely small beyond $n/2$. We used $|\mathcal{S}| = 40$, but for larger datasets, it can be set to be larger if computationally feasible while also considering the approximation/speed trade-off. Lastly, we used $\lambda \in \{1, 20, 50\}$.

Results. We show the prediction accuracies of our algorithm in Table 1. We run the classification tasks for $K \in \{2, 3, 4\}$ respectively. We compare the similarities of the bases of each test subject (using ℓ_2 norm) to the set of bases in each partition to locate the closest one, which is the assigned quantile label of the test subject. In Table 1, we show the accuracy results for RAVLT and MMSE using four setups (columns 2 through 5): **(a)** non-coupled, **(b)** longitudinally coupled, **(c)** cross-sectionally coupled and **(d)** both

K	Non-coupled		Longitudinal		Cross-section		Coupled	
	$j=1$	$\{1, 2, 3\}$	1	$\{1, 2, 3\}$	1	$\{1, 2, 3\}$	1	$\{1, 2, 3\}$
R:2	33.33	34.92	42.86	42.86	66.67	60.32	71.43	71.43
R:3	38.10	33.33	52.38	36.51	57.14	44.44	57.14	55.56
R:4	28.57	28.57	23.81	30.16	30.16	23.81	47.62	34.92
M:2	42.86	41.27	28.57	30.16	57.14	39.68	76.19	71.43
M:3	42.86	38.10	47.62	49.21	47.62	46.03	47.62	50.79
M:4	34.92	28.57	23.81	14.29	19.05	12.70	47.62	28.57

Table 1: Prediction accuracy (%) of RAVLT (R: $K \in \{2, 3, 4\}$ quantiles) and MMSE (M: $K \in \{2, 3, 4\}$ quantiles) on $j = 1$ time point and $j = \{1, 2, 3\}$. Best results are in red.

longitudinally and cross-sectionally coupled. For RAVLT, first, we discuss the simplest setting for $K = 2$ (R:2 in Table 1). Here, the performance estimates suggest that the first three setups are unable to identify the signal whereas our proposed coupled setup offers accuracy estimates approaching 70%. This trend of the coupled setup improving the accuracy of the non-coupled or partially coupled setup continues for $K = 3$ (R:3) and $K = 4$ (R:4). We observe a very similar trend for the MMSE quantile prediction task suggesting that capturing the full set of longitudinal data in terms of its coupled bases offers significant advantages for predicting subject-level cognitive status.

Discussion. We briefly elaborate on the relevance of these findings. Recall that our dataset is preclinical, i.e., all subjects are healthy. This means that the brain connectivity changes that we seek to capture using our proposed formulation are *extremely subtle*. To appreciate the small effect sizes in this dataset, we show in Fig. 5 the actual brain connectivity adjacency matrices for $K = 3$ from the RAVLT based staging. The numbers at the top of each matrix image is the *total* sum of edge weights in the graph. As expected, when we move from left to right (and top to bottom), the overall connectivity progressively becomes weaker but the changes are *extremely small* and nearly impossible to pick up in a statistically significant way if this analysis were conducted on an edge-by-edge manner. Despite the fact that the overall changes are in the 0.5-1.0% range over the entire set of 12720 edges, our coupled harmonic bases setup is still able to offer better than chance prediction accuracy. Achieving this capability in a preclinical population is the main scientific result of our experimental evaluations.

Finally, we note that our model recovers disease effects at the level of individual tracts reliably: it specifically identified the top 50 fiber tracts with the most changes in connectivity strength across RAVLT progression (illustrative results are described in detail in the supplement).

5. Conclusion

The goal of this paper is to characterize the evolutionary patterns of brain connectivity networks derived from a longitudinal set of middle-aged healthy individuals who are at risk for Alzheimer’s disease. The changes in brain connec-

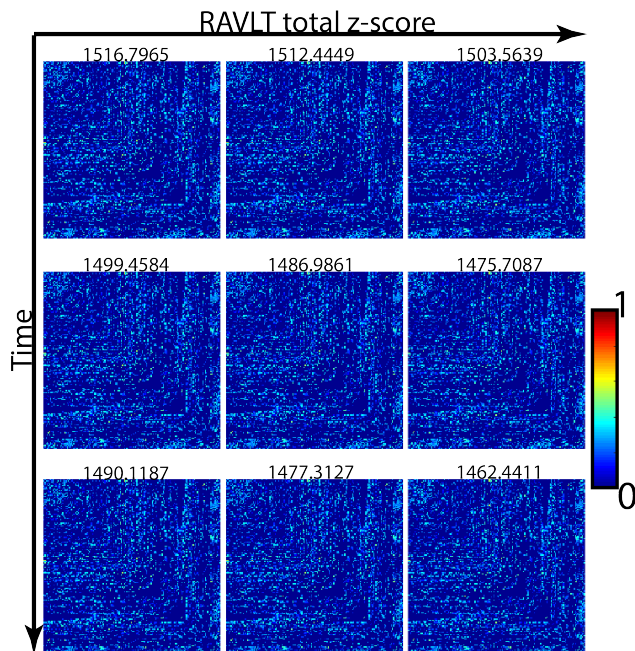


Figure 5: Average adjacency matrices for each of the three stages (columns) based on RAVLT total z-scores and each of the three time points (rows). The total magnitude of connectivity strengths (sum of the total edge weights) are shown in the respective titles of the matrices. Even though there is a trend of decrease in the overall connectivity strength along the cognitive staging, the effect sizes are extremely small 0.5 – 1.0%. In the case of individual edges the effects are even smaller.

tivity are extremely small during the preclinical stages and existing approaches do not seem to be sensitive enough. We presented a framework which treats the entire set of graph Laplacians of the brain connectivity networks as a system by explicitly considering the coupling between different cognitive as well as the longitudinal (temporal) stages. Our experimental results provide evidence that such a coupled bases approach can indeed provide better insights into the brain network changes across the clinical stages. While the technical development of our framework was motivated by the neuroimaging application, the resultant numerical optimization schemes can be widely applicable for incorporating relevant couplings into generalized eigenvalue problems which are pervasive in many other areas of computer vision and machine learning. The extended version of this paper, the supplement and the code are available at <http://pages.cs.wisc.edu/~sjh/>.

6. Acknowledgment

SJH was supported by a University of Wisconsin CIBM fellowship (5T15LM007359-14). We acknowledge support from NIH R01 AG040396 (VS), NIH R01 AG027161 (SCJ), NIH R01 AG37639 (BBB), NSF CCF 1320755 (VS), NSF CAREER award 1252725 (VS), UW ADRC AG033514, UW ICTR 1U1RR025011, Waisman Core grant P30 HD003352-45 and UW CPCP A1117924.

References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. 4, 5
- [2] S. Achard, R. Salvador, B. Whitcher, et al. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of neuroscience*, 26(1):63–72, 2006. 2
- [3] N. J. Cairns, R. J. Perrin, E. E. Franklin, et al. Neuropathologic assessment of participants in two multi-center longitudinal observational studies: The Alzheimer Disease Neuroimaging Initiative (ADNI) and the Dominantly Inherited Alzheimer Network (DIAN). *Neuropathology*, 2015. 1
- [4] M. D. Collins, J. Liu, J. Xu, et al. Spectral clustering with a convex regularizer on millions of images. In *ECCV*. 2014. 5
- [5] P. Cook, Y. Bai, S. Nedjati-Gilani, et al. Camino: open-source diffusion-MRI reconstruction and processing. In *ISMRM*, volume 2759, 2006. 7
- [6] A. Drzezga, J. A. Becker, K. R. Van Dijk, et al. Neuronal dysfunction and disconnection of cortical hubs in nondemented subjects with elevated amyloid burden. *Brain*, 134(6):1635–1646, 2011. 3
- [7] J. Ham, D. Lee, and L. Saul. Semisupervised alignment of manifolds. In *UAI*, volume 10, pages 120–127, 2005. 3
- [8] X. Hua, A. D. Leow, N. Parikshak, et al. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer’s disease: an MRI study of 676 AD, MCI, and normal subjects. *NeuroImage*, 43(3):458–469, 2008. 6
- [9] S. J. Hwang, M. D. Collins, S. N. Ravi, et al. A Projection Free Method for Generalized Eigenvalue Problem With a Nonsmooth Regularizer. In *ICCV*, 2015. 5
- [10] S. Jbabdi, S. N. Sotiropoulos, S. N. Haber, et al. Measuring macroscopic brain connections in vivo. *Nature neuroscience*, 18(11):1546–1555, 2015. 1
- [11] S. Keihaninejad, H. Zhang, N. S. Ryan, et al. An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer’s disease. *NeuroImage*, 72:153–163, 2013. 6
- [12] W. H. Kim, N. Adluru, M. K. Chung, et al. Multi-resolutional brain network filtering and analysis via wavelets on non-Euclidean space. In *MICCAI*. Springer, 2013. 1
- [13] W. H. Kim, N. Adluru, M. K. Chung, et al. Multi-resolution statistical analysis of brain connectivity graphs in preclinical Alzheimer’s disease. *NeuroImage*, 118:103–117, 2015. 1
- [14] A. Kovnatsky, M. M. Bronstein, A. M. Bronstein, et al. Coupled quasi-harmonic bases. In *Computer Graphics Forum*, volume 32, pages 439–448. Wiley Online Library, 2013. 2
- [15] Z. Lei and S. Z. Li. Coupled spectral regression for matching heterogeneous faces. In *CVPR*, 2009. 2
- [16] J. Liu, S. J. Wright, C. Ré, et al. An asynchronous parallel stochastic coordinate descent algorithm. *JMLR*, 16(1):285–322, 2015. 6
- [17] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006. 5
- [18] A. Raj, A. Kuceyeski, and M. Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215, 2012. 1
- [19] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, et al. The application of graph theoretical analysis to complex networks in the brain. *Clinical Neurophysiology*, 118(11):2317–2331, 2007. 2
- [20] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010. 2
- [21] Y. Saad. *Numerical methods for large eigenvalue problems*, volume 158. SIAM, 1992. 2
- [22] M. Schmidt et al. *Key auditory verbal learning test: a handbook*. Western Psychological Services Los Angeles, 1996. 7
- [23] S. N. Sotiropoulos, S. Jbabdi, J. Xu, et al. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*, 80:125–143, 2013. 1
- [24] O. Sporns. The human connectome: a complex network. *Annals of the New York Academy of Sciences*, 1224(1):109–125, 2011. 1
- [25] C. J. Stam and J. C. Reijneveld. Graph theoretical analysis of complex networks in the brain. *Nonlinear biomedical physics*, 1(1):3, 2007. 2
- [26] T. N. Tombaugh and N. J. McIntyre. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, 40(9):922–935, 1992. 7
- [27] K. Uğurbil, J. Xu, E. J. Auerbach, et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*, 80:80–104, 2013. 1
- [28] D. C. Van Essen, K. Ugurbil, E. Auerbach, et al. The Human Connectome Project: a data acquisition perspective. *NeuroImage*, 62(4):2222–2231, 2012. 1
- [29] A. Varentsova, S. Zhang, and K. Arfanakis. Development of a high angular resolution diffusion imaging human brain template. *NeuroImage*, 91:177–186, 2014. 7
- [30] C. Wang and S. Mahadevan. Manifold alignment using Procrustes analysis. In *ICML*, 2008. 3
- [31] P. Wang, B. Zhou, H. Yao, et al. Aberrant intra-and inter-network connectivity architectures in Alzheimer’s disease and mild cognitive impairment. *Scientific reports*, 5, 2015. 1
- [32] J. S. Warsa, T. A. Wareing, J. E. Morel, et al. Krylov subspace iterations for deterministic k-eigenvalue calculations. *Nuclear Science and Engineering*, 147(1):26–42, 2004. 2
- [33] Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1-2):397–434, 2013. 5, 6
- [34] G. S. Wig, B. L. Schlaggar, and S. E. Petersen. Concepts and principles in the analysis of brain networks. *Annals of the New York Academy of Sciences*, 1224(1):126–146, 2011. 1
- [35] H. Zhang, B. B. Avants, P. Yushkevich, et al. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *TMI*, 26(11):1585–1597, 2007. 7