Matched Signal Detection on Graphs: Theory and Application to Brain Network Classification

Chenhui Hu[1,2], Lin Cheng[3], Jorge Sepulcre[1], Georges El Fakhri[1], Yue M. Lu[2], and Quanzheng Li[1]

[1]Center for Advanced Medical Imaging Science, NMMI, Radiology, Massachusetts General Hospital, Boston, MA 02114
{sepulcre@nmr.,elfakhri@pet.,Li.Quanzheng@}mgh.harvard.edu
[2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138
{hu4,yuelu}@seas.harvard.edu
[3]Department of Engineering, Trinity College, Hartford, CT 06106
lin.cheng@trincoll.edu

Abstract. We develop a matched signal detection (MSD) theory for signals with an intrinsic structure described by a weighted graph. Hypothesis tests are formulated under different signal models. In the simplest scenario, we assume that the signal is deterministic with noise in a subspace spanned by a subset of eigenvectors of the graph Laplacian. The conventional matched subspace detection can be easily extended to this case. Furthermore, we study signals with certain level of smoothness. The test turns out to be a weighted energy detector, when the noise variance is negligible. More generally, we presume that the signal follows a prior distribution, which could be learnt from training data. The test statistic is then the difference of signal variations on associated graph structures, if an Ising model is adopted. Effectiveness of the MSD on graph is evaluated both by simulation and real data. We apply it to the network classification problem of Alzheimer's disease (AD) particularly. The preliminary results demonstrate that our approach is able to exploit the sub-manifold structure of the data, and therefore achieve a better performance than the traditional principle component analysis (PCA).

Keywords: Matched subspace detection, graph-structured data, graph Laplacian, brain networks, classification, Alzheimer's disease

1 Introduction

Matched subspace detection is a classic tool that determines whether a multidimensional signal lies in a given linear subspace or not [1]. It has achieved a great success in applications such as radar, hyperspectral imaging and medical imaging [2]. The subspace is either governed by the physical system that generates the signal, or could be inferred from training data. Subspace learning

or dimensionality reduction is a central issue in machine learning. One popular method is principal component analysis (PCA), which projects the original data to a linear subspace spanned by the leading eigenvectors of the data matrix. A common assumption of PCA is that the data are generated from a linear subspace. In fact, many real data are sampled from a nonlinear low-dimensional *sub-manifold*, which is embedded in a high-dimensional ambient space [3]. Examples include images, gene data, social network records, and sensor network measurements. These types of data may be better modeled as signals supported on graph structures,[1] instead of conventional signals in Euclidean spaces. In this setting, the signal subspace can be effectively learnt by graph spectral methods, *e.g.*, Isomap, Locality Linear Embedding (LLE), Laplacian eigenmaps [4, 5].

Motivated by the requirement of classifying graph-structured data in many emerging problems, we are interested in developing a similar detection framework for graph-signals in this paper. Specifically, we formulate hypothesis tests to decide which graph structure a signal is more likely to embed in. Instead of building combinatorial tests on graph [6, 7], we exploit the matched subspace detection technique to make our setup generic to handle a variety of situations. Intuitively, the tests are dependent on the relation between the signal and the graphs, *i.e.*, the signal models. To this end, we first assume that the signal lies in a subspace spanned by a subset of eigenvectors of the graph Laplacian matrix. The classic matched subspace detection can be applied directly. Then, we consider signals that are *smooth* on graph, as specified by a bounded variation metric. The maximum likelihood estimator (MLE) of the true signal is derived by solving a constrained optimization problem. When the noise variance is negligible, we find the test becomes a weighted energy detector. More generally, we presume the signal is randomly drawn from a prior distribution. It ends up with comparing the signal variations on the hypothetic graphs, if an Ising model is adopted.

We apply the proposed detection theory to brain network classification for AD. As one of the most common forms of dementia, AD is believed to be a brain network disease, and is characterized by progressive impairment of memory and other cognitive capacity, which eventually causes death. It affects nearly 36 million people in the world with an expected number 65.7 million by 2030 [8]. While conventional clinical diagnosis might be inaccurate, Positron Emission Tomography (PET) imaging of brain amyloid using Pittsburgh Compound-B (PIB) tracer provides sensitive and consistent biomarkers in the early stage of the disease [9]. We carry out leave-one-out tests on 30 AD patients and 40 normal control (NC) subjects. Experimental results show that when using the MSD on graph, the probabilities of false alarm and miss are 2/40 and 0/30, respectively. In contrast, the associated probabilities are 6/40 and 5/30, if a linear PCA is used; or 5/40 and 3/30, if we use support vector machine (SVM). This preliminary result indicates the MSD on graph provides an effective way for AD network classification, probably due to the effectiveness of exploiting the sub-manifold structure of the data.

[1] For short, we will also refer to them as graph-signals.

2 Signal Models

Since we aim to classify graph-structured data, such as neuroimaging data, it is necessary to introduce the statistical signal models on graph before presenting the hypothesis testing models. A brain network can be represented by a *weighted graph* $\mathcal{G}(\mathcal{V}, \mathcal{E}, W)$ with vertex set $\mathcal{V}(|\mathcal{V}| = N)$ and edge set $\mathcal{E}$. For such graph, the similarity or associativity between vertex i and j is given by $W_{ij} \geq 0$. Besides, we introduce a diagonal degree matrix D with $D_{ii} = \sum_{j=1}^{N} W_{ij}$. Then, the *Laplacian matrix* of the graph is defined as $L \stackrel{\text{def}}{=} D - W$. Note that L is symmetric, we can decompose it into $L = F\Lambda F^T$, where Λ is diagonal with $\Lambda_{ii} = \lambda_i$ being the i-th smallest eigenvalue of L and the columns of F, f_is, are the associated eigenvectors. By nature, the eigenvalues of L satisfy: $0 = \lambda_1 \leq \lambda_2 \cdots \leq \lambda_N$.

Let $\mathcal{H}$ be a Hilbert space on $\mathcal{V}$. A signal x on graph $\mathcal{G}$ is a $N \times 1$ vector in $\mathcal{H}$, with each entry being a real value assigned to a vertex (see Fig. 1 for an example). Since F is orthogonal, the projection of x onto its column space is

$$\widehat{x} = F^T x. \tag{1}$$

Accordingly, we have $x = F\widehat{x}$. We refer to $\widehat{x}$ as the *graph Fourier transform* (GFT) of x and x the inverse GFT of $\widehat{x}$ [10, 11], based on the fundamental connection between (1) and the classical discrete Fourier transform (DFT): the eigenvectors of the Laplacian matrice form a DFT basis, for any circulant graph.

To facilitate the proposed hypothesis tests (whether a signal is embedded in a given graph structure), prior information could be applied. It could simply be additional constraints. Alternatively, we can assume that the signal follows a prior distribution, which could be learnt from the training data. Next, we illustrate the signals with both types of prior information in this section.

2.1 Finite Support Graph-Signals

In many applications, the GFT component of the signal is more likely to be close to zero as the corresponding eigenvalue becomes large. This is largely due to the *smoothness* of the signal on graph, since a higher eigenvalue reflects a stronger variation of the eigenvector [12, 13]. Analogous to the traditional signal, we call the f_is frequency components for GFT. Generally, we can define signals that are only supported on selected frequency components, namely the *finite support signals*, as specified by $\left\{x \mid \widehat{x}_{i \notin I_S} = 0, I_S \subset \{1, \cdots, N\}\right\}$. Usually, the supporting eigenvectors in I_S are chosen on the basis of the amplitudes of the GFT components. By introducing a binary diagonal matrix S with $S_{ii} = 1$ only if $\widehat{x}_i \neq 0$, we can write $x = FS\widehat{x}$.

2.2 Constrained Graph-Signals

To improve the robustness of a hypothesis test, prior information is usually applied. One common prior is to add regularization to the GFT coefficients in

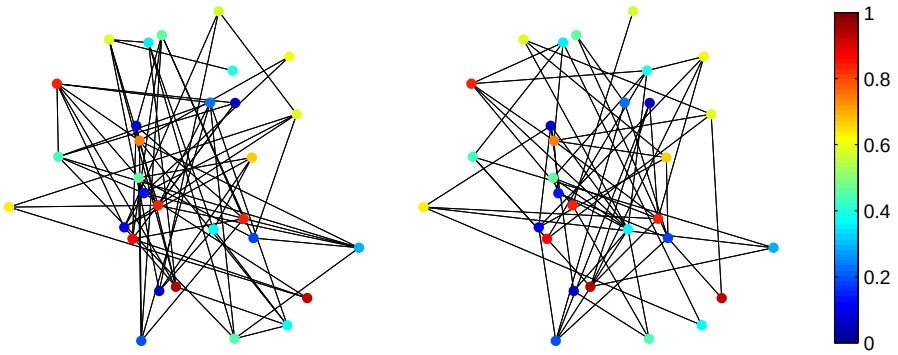


Fig. 1. A same graph-signal displayed on two graph structures. Both graphs have 30 vertices but different connections, each of which has a unit weight. The values of the signal on the vertices are encoded in the color of the dots.

the form of a penalty function $C(\cdot)$, which imposes a reward or penalty when the signal is in a certain shape. Due to the uncertainty caused by the noise, the hypothesis tests involve the estimation of maximum likelihood (ML) with unknown parameters. Under the constraint, we maximize the following penalized ML instead of ML:

$$\arg\max_{\theta} LL(\widehat{x}, \theta_{\setminus \widehat{x}}) - \gamma \cdot C(\widehat{x}), \tag{2}$$

where $LL(\cdot)$ is the log-likelihood function, $\gamma > 0$ is a trade-off factor, and $\theta = \{\widehat{x}, \theta_{\setminus \widehat{x}}\}$ denotes the set of all unknown but *deterministic* parameters. In Section 3.2, we replace the corresponding parameter with its penalized MLE and obtain a general likelihood ratio (GLR) test.

2.3 Probabilistic Graph-Signals

Perviously, we treat the signal as deterministic but unknown due to noise. From a Bayesian perspective, we can also assume that the signal is randomly generated from a prior distribution.

The distribution could be expressed in the graph domain. In [15, 16], the authors proposed a so-called *Ising model* as follows

$$\mathbb{P}(x) \propto \exp(-x^T L x). \tag{3}$$

The exponent term in the above equation is the variation of x on the graph for that $x^T L x = \sum_{i,j} W_{ij}(x_i - x_j)^2$ [12]. PIB-PET images for a certain subject group may obey this model, since the correlated brain regions should yield similar measurements. In the GFT domain, we consider a multivariate Gaussian distribution of the GFT coefficients. We will define a probabilistic MSD on graph in Section 3.3 based on above probabilistic graph-signals.

3 Decision Models

Let $y \in \mathbb{R}^N$ be an observed signal defined on a potential graph. There are two hypotheses $H_j, j = 0, 1$, representing that y is embedded in either graph $\mathcal{G}_0$ or $\mathcal{G}_1$, with the associated Laplacian matrix L_0 or L_1, respectively. The graphs are defined on the same set of vertices. In Fig. 1, we show a same graph-signal on two different hypothetic graphs, where every edge has a unit weight. Our goal is to formulate a hypothesis test to decide which graph fits the signal more accurately. We refer to the procedure as *matched signal detection* (MSD) on graph.

We first presume that y is a contaminated version of a signal x with additive Gaussian noise, since the observed signal may not fit the learnt model completely. Namely, we express $y = x + n$, with $n \sim \mathcal{N}(\mathbf{0}, \Phi)$. Without loss of generality, we assume $\Phi = \xi^2 I$. Followed from (1), we have

$$y = F\widehat{x} + n. \tag{4}$$

Because the observation y is random, the most general form of the likelihood ratio (LR) test would be

$$\frac{l(\theta_1; y)}{l(\theta_0; y)} = \frac{\mathbb{P}(y|H_1)}{\mathbb{P}(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1} = \eta, \tag{5}$$

where $l(\cdot)$ indicates the likelihood function, θ_j and π_j are the set of parameters and prior probability under H_j, respectively. By default, we choose equal priors, *i.e.*, $\eta = 1$. In accordance with the signal models in last section, we present more concrete versions of (5) in terms of different types of MSD on graph.

3.1 Simple MSD on Graph

The basic form of MSD on graph follows immediately from the conventional matched subspace detection, if the subspace is spanned by a subset of the eigenvectors of the Laplacian matrix under each hypothesis. Suppose the eigendecomposition of L_j possesses the form $L_j = F_j \Lambda_j F_j^T$, from the orthogonality of F_j, we could obtain the test statistic

$$T_1(y) = \frac{\|(I - S)F_0 y\|_2^2}{\|(I - S)F_1 y\|_2^2}, \tag{6}$$

where S is the indicative matrix of the GFT support defined in Section 2.1.

3.2 CMSD on Graph

We adopt CMSD as a shorthand for the constrained MSD problem, when the graph-signal model with constraint in Section 2.2 is taken into account. By multiplying F^T to both sides of (4), we obtain

$$\widehat{y} = \widehat{x} + \widehat{n}, \tag{7}$$

where $\widehat{\cdot}$ denotes the GFT of the corresponding parameter. Since F is orthonormal, $\widehat{n}$ is still a white Gaussian noise of the same distribution as n.

The constraint function may be selected in various ways. For instance, if we prefer a sparse $\widehat{x}$ as that in [17], we can let $C(\widehat{x}) = \|\widehat{x}\|_1$. Here we introduce the following quadratic penalty function

$$C(\widehat{x}) = \sum\nolimits_{i\in I_S} \alpha_i \widehat{x}_i^2, \tag{8}$$

with α_i being a non-negative weight. We might assign a larger penalty weight on the GFT coefficient which is less informative. For the CMSD with above quadratic constraint, we have the following theorem

Theorem 1 *Given the cost function (8) under two hypotheses and denote by $\widehat{y}_{i,j}$ the i-th entry of $F_j^T y$, the GLRT statistic can be expressed as*

$$T_2(y) = \frac{\sum_{i\in I_S} \alpha_i \widehat{y}_{i,0}^2}{\sum_{i\in I_S} \alpha_i \widehat{y}_{i,1}^2}, \tag{9}$$

if the noise variances are unknown but significantly small.

Proof. See Appendix A.

We then present a specific form of CMSD by imposing a smoothness constraint to the graph-signals measured on the graph structure. This kind of constraint is common and is particularly suitable for neuroimaging data, since brain imaging data generated from similar physiological mechanisms would vary smoothly along a certain sub-manifold structure. To measure the degree of smoothness, for $s > 0$, we define the following metric on graph for

$$V_{\mathcal{G},s}(x) = \frac{x^T L^s x}{x^T x}. \tag{10}$$

Lemma 1 *(1) $x^T L^s x = \sum_i \lambda_i^s \widehat{x}_i^2$; (2) $V_{\mathcal{G},1}(f_i) = \lambda_i$,*

The proof is omitted due to limitation of space. Lemma 1 indicates the smoothness measurement is a special case of the constraint in (8). From the lemma, we observe that the signal will have a large variation when its GFT components are concentrated on high frequency components, *i.e.*, f_i with large λ_i.

We consider the smoothness constraint as a *bounded variation* of the signal. Namely, we assume $V_{\mathcal{G},s}(x) < r$ holds for $0 < r < \lambda_N$. In particular, when $s = 1$, we obtain that $\sum_{i=1}^N \lambda_i \widehat{x}_i^2 < r \sum_{i=1}^N \widehat{x}_i^2$, *i.e.*, $\sum_{i=1}^N (\lambda_i - r)\widehat{x}_i^2 < 0$. Under this constraint and by the KKT condition, we have

$$\frac{\partial}{\partial \hat{x}} \Big\{ N \log \xi + \frac{1}{2\xi^2} \|\widehat{x} - \widehat{y}\|_2^2 + \gamma \|R\widehat{x}\|_2^2 \Big\} = 0, \tag{11}$$

$$\gamma \big(\sum_{i=1}^N (\lambda_i - r)\widehat{x}_i^2 \big) = 0, \quad \gamma \geq 0. \tag{12}$$

Notice that R is a diagonal matrix with $R_{ii} = (\lambda_i - r)^{1/2}$ and γ is a dual variable. Following the proof of Theorem 1, we get $\widehat{x}_i = \frac{\widehat{y}_i}{1+\gamma\xi^2(\lambda_i - r)}$ from (11). Replacing $\widehat{x}_i$ in (12) with this, we can write

$$\sum_{i<\tau} \frac{(r-\lambda_i)\widehat{y}_i^2}{(1+\gamma\xi^2(\lambda_i - r))^2} = \sum_{i>\tau} \frac{(\lambda_i - r)\widehat{y}_i^2}{(1+\gamma\xi^2(\lambda_i - r))^2}, \tag{13}$$

where $\tau = \max\{i|\lambda_i < r\} + \frac{1}{2}$. If we assume that the noise variance ξ can be estimated from the average residual energy of the projection as

$$\xi = \sqrt{\frac{\sum_{i=N'+1}^{N} \widehat{y}_i^2}{N - N'}}, \tag{14}$$

with $N' < N$ being an integer threshold, then from (13) we can solve out the dual parameter γ. After that, the MLE of $\widehat{x}$ can be readily obtained. Plugging it to the likelihood expression in (5), we will reach the GLRT.

3.3 PMSD on Graph

A more general MSD is to consider random graph-signals, which gives rise to the probabilistic MSD (PMSD) on graph. We assume that the p.d.f. of $\widehat{x}$ is $g_j(\widehat{x})$ under H_j. By independence of the noise on the signal, the LR is

$$LR = \frac{\int g_1(\widehat{y} - t)h_1(t)dt}{\int g_0(\widehat{y} - t)h_0(t)dt}, \tag{15}$$

where $h_j(\cdot)$ denotes the p.d.f. of the white Gaussian noise under H_j.

If the GFT coefficients of the true signal follow a Gaussian distribution $\mathcal{N}(\mathbf{0}, \Sigma_j)$, the test statistic would be $\widehat{y}^T(\Phi_0^{-1} - \Phi_1^{-1})\widehat{y}$, where $\Phi_j = \Sigma_j + \xi_j^2 I$. In particular, if we apply the Ising model (3) to the true signal, we will have $g_j(\widehat{x}) \propto e^{-(F_j\hat{x})^T L_j (F_j\hat{x})} = e^{-\hat{x}^T \Lambda_j \hat{x}}$. It turns out that $\widehat{y}$ is distributed as $\mathcal{N}(\mathbf{0}, \Lambda_j^{\dagger} + \xi_j^2 I)$, by viewing g_j as a degenerated Gaussian distribution. Here $\Lambda_j^{\dagger}$ is the pseudoinverse of Λ. When the noise variances are known, the LRT statistic reduces to

$$T_3(y) = \sum_i \beta_{i,0}\widehat{y}_{i,0}^2 - \sum_i \beta_{i,1}\widehat{y}_{i,1}^2, \tag{16}$$

with $\beta_{1,j} = \xi_j^{-2}$ and $\beta_{i,j} = (\lambda_{i,j}^{-1} + \xi_j^2)^{-1}$, for $i \geq 2$. In a special noise-free case, the statistic in (16) becomes $y^T(L_0 - L_1)y$, which is simply a measure of the difference of the signal variations on the two graph structures.

4 Experimental Validation

4.1 Numerical Simulation

In this section, we evaluate the MSD on a pair of small-world networks. It has been reported that human brain networks have a small-world topology that supports both segregated and distributed information processing with a minimized

cost [18]. The alternation of the graph structures is often caused by certain diseases. Previously, people used graph metrics, like degree distribution, average shortest path length, clustering coefficient, *etc.*, to classify different networks. Here we fulfill the task upon two networks characterized by the Watts-Strogatz model [19], using the MSD schemes developed in Section 3.

The networks are constructed from random rewiring of a circulant graph with 40 vertices. The number of edges of each vertex and the rewiring probability are (12, 0.05) under H_0, and (20, 0.5) under H_1, respectively. We assume that the GFT coefficients of the signals are distributed as $\mathcal{N}(\mathbf{0}, \Sigma)$ with Σ being diagonal and $\Sigma_{ii} = \exp(-i/10)$ for both hypotheses. The standard deviation of the noise $\xi = 0.1$. In the first case, we presume that the signal distribution and

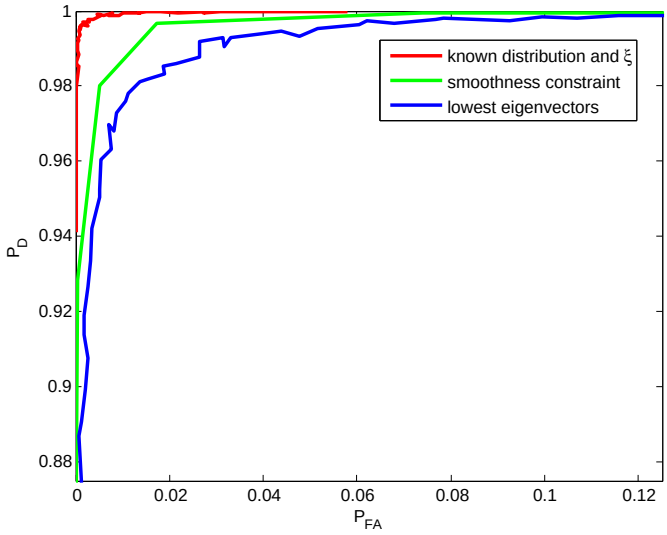


Fig. 2. ROC curves for hypothesis tests upon small-world networks of 40 vertices with vertex degree and rewiring probability being $(12, 0.05)$ and $(20, 0.5)$, respectively.

noise level are known. As shown in Fig. 2, we can apply the PMSD to get an almost perfect ROC curve in red. In the second case, we impose a smoothness constraint specified by an upper bound equal to λ_{20} on the signal variation (10) in each graph. At last, we implement the simple MSD via picking up the first 20 eigenvectors. The green and the blue lines display the associated ROC curves for them. We find that the CMSD can outperform the simple MSD and is slightly inferior to the optimal case. Notice that the areas under the ROC curves are all close to 1, verifying the effectiveness of the proposed approaches.

4.2 AD Network Classification

We then apply the proposed MSD to the classification of PIB-PET images and compare it with PCA in this section. The data set consists of 30 AD patients and 40 Normal Control (NC). Part of the data has been studied in [20] as well. Among the normal subjects, 20 are labeled as PIB positive and the rest as PIB negative based on the beta-amyloid binding level. Each image has a dimension of $20 \times 24 \times 18$ with $8mm \times 8mm \times 8mm$ voxels, downsampled from $2mm \times 2mm \times 2mm$ voxels for computational efficiency. In the pre-processing of the data, we first mask out the area out of the brain. Next, we apply Automated Anatomical Labeling (AAL) [21] to map the effective voxels to 116 volumes-of-interest (VOIs). The data is then averaged within each VOI for further analysis.

As an initial step, we build a similarity graph over 42 regions among all the VOIs that are regarded to be potentially related to AD. Table 1 in [22] lists the names of the VOIs spread over the frontal, parietal, occipital, and temporal lobes. For an arbitrary group, let $\{R_1, \cdots, R_p\}$ be the p selected VOIs and suppose we have m samples. The observation in the i-th region of subject j is denoted by x_{ij}. We construct a weighted graph over the p brain volumes, by assigning a positive weight W_{ij} to the edge between R_i and R_j as follows

$$W_{ij} = \exp\left(- \frac{\|x_i - x_j\|^2}{\rho^2} \right), \tag{17}$$

where $x_i = (x_{i1}, \cdots, x_{im})^T$, $x_j = (x_{j1}, \cdots, x_{jm})^T$, and $\rho > 0$ is a scaling factor (here we set $\rho = 1$). The kernel function in (17) is known as a heat kernel [3]. After building graphs for both groups, we project a newly observed signal to the sets of eigenvectors of the graph Laplacian matrices. The decision is made through comparing the test statistic (6) against one. We present the major steps of our data processing in Fig. 3.

We perform leave-one-out tests to evaluate the proposed MSD on graphs. Fig. 4 demonstrates the projection errors, *i.e.*, the numerator and denominator in (6), when the true signal belongs to NC and AD, respectively. Here we merely choose the first 4 eigenvectors as the supporting set. The error rates in Fig. 4(a) and 4(b) are 2/40 and 0/30, compared with the rates 6/40 and 5/30 when we use linear PCA, namely we form a matrix by aligning the existing data and use the principle left eigenvectors of the data matrix to carry out the matched subspace detection. We've also carried out the classification by SVM [23], which gives the probabilities of false alarm and miss 5/40 and 3/30. It shows the advantage of the MSD on graph over the traditional method and indicates that the data could have an intrinsic sub-manifold structure. We also observe that there are less differences of projection errors in NC than in AD. This might be due to the following reasons: (1) The PIB binding levels of NC subjects are about 20% lower than those of the AD subjects on average; (2) We reduce the dimension of the raw data by mapping it to a few VOIs, indicating some information may not be retained; (3) We use a simple strategy to construct the weighted graphs which could be improved via more sophisticated schemes (*e.g.*, [24]). This observation also implies that if we slightly increase the threshold of our hypothesis test,

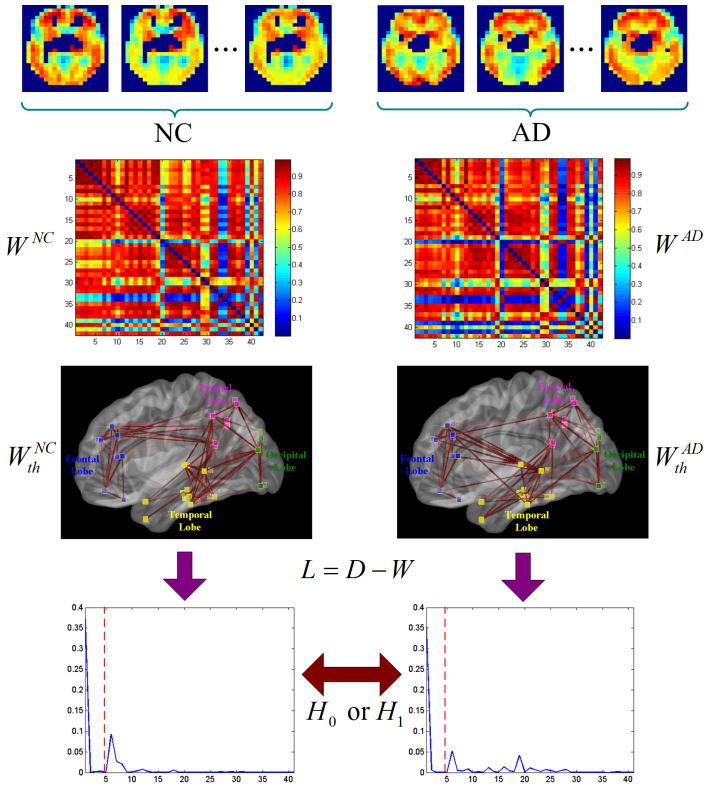
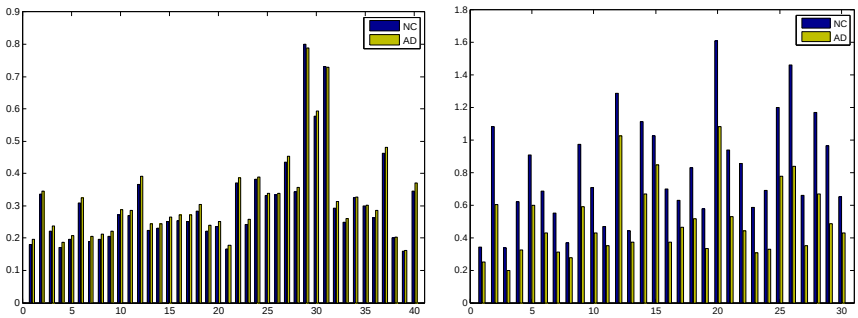


Fig. 3. Main steps (from the top to the bottom) of the MSD implementation for AD network classification. W^{NC}, W^{AD} are the weighted graphs constructed from the imaging data listed in the first row; while W_{th}^{NC}, W_{th}^{AD} are their corresponding thresholded versions with 135 edges. The last row illustrates the projection energy distributions of a signal on the two sets of eigenvectors of the graph Laplacians.



(a) NC group (b) AD group

Fig. 4. Projection errors in the leave-one-out tests when the true data is from either NC or AD group. Note that the x-axis is the index of subjects.

we can achieve 100% detection rate. In addition, we find that the two miss-classified subjects (Number 29 and 31) in Fig. 4(a) are both PIB positive, which may indicate that they are more likely to develop AD.

5 Conclusion

In this paper, we formulate the MSD for graph-structure data under different signal models. We consider the signals that are either smooth on the graph or randomly generated from a known prior distribution. In the first setting, GLRT can be obtained by solving a constrained optimization problem. Specially, when the noise variance is negligible, it results in a weighted energy detector. In the second case, the test statistic is the difference of the signal variations on the graphs, if an Ising model is employed. We test the effectiveness of the MSD on simulated and real data by applying it to AD network classification. Compared with the linear PCA, our method demonstrates a better performance due to its ability of exploiting the sub-manifold structure of the neuroimaging data.

Appendix A: Proof of Theorem 1

Define a diagonal matrix R as such $R_{ii} = \alpha_i^{1/2}$ for $i \in I_S$, and $R_{ii} = 0$ elsewhere, then the penalty function can be written as $C(\widehat{x}) = \|R\widehat{x}\|_2^2$. We need to estimate the set of parameters $\theta = \{\widehat{x}, \xi\}$ from the following

$$\arg\min_{\hat{x},\xi} N \log \xi + \frac{1}{2\xi^2}\|\widehat{x} - \widehat{y}\|_2^2 + \gamma\|R\widehat{x}\|_2^2, \tag{18}$$

when the noise variance is unknown. Denote by $Q(\widehat{x}, \xi)$ the objective function in the above equation and set $\frac{\partial Q}{\partial \widehat{x}} = 0$, we obtain $(\widehat{x} - \widehat{y})/\xi^2 + \gamma R^2\widehat{x} = 0$, *i.e.*, $\widehat{x} = (I + \xi^2\gamma R^2)^{-1}\widehat{y}$. Similarly, let $\frac{\partial Q}{\partial \xi} = 0$, we have $\xi^2 = \|\widehat{x} - \widehat{y}\|_2^2 = \|(I+\xi^2\gamma R^2)^{-1}\widehat{y} - \widehat{y}\|_2^2$. If $\xi \ll \frac{1}{\gamma \max_i\{\alpha_i\}}$, we can proceed with the approximation $\xi^2 = \|\xi^2\gamma R^2\widehat{y}\|_2^2$, indicating that $\xi = \gamma^{-1}\|R\widehat{y}\|_2$. Plugging it into (5) leads to the constrained test here.

References

1. Scharf, L.L., Friedlander, B.: Matched Subspace Detectors. IEEE Trans. on Signal Proc. 42, 2146–2157 (1994)
2. Li, Z., Li, Q., Yu, X., Conti, P.S., Leahy, R.M.: Lesion Detection in Dynamic FDG-PEG Using Matched Subspace Detection. IEEE Trans. on Medical Imaging. 28, 230–240 (2009)
3. Belkin, M., Niyogi, P.: Using Manifold Structure for Partially Labeled Classification. Advances in Neural Information Processing Systems. 15, 929–936 (2002)
4. Saul, L.K., Weinberger, K.Q., Ham, J.H., Sha, F., Lee, D.D.: Spectral Methods for Dimensionality Reduction. Semisupervised learning, In: O. Chapelle, B. Schoelkopf, and A. Zien (eds.), Semisupervised Learning, pages 293-308. MIT Press: Cambridge, MA (2006)

5. Cai, D., He, X., Han, J.: Spectral regression: A Unified Approach for Sparse Subspace Learning. In: 7th IEEE Int. Conf. on Data Mining, pp. 73–82. (2007)
6. Addario-Berry, L., Broutin, N., Devroye, L., Lugosi, G.: On Combinatorial Testing Problems. The Annals of Statistics. 38, 3063–3092 (2010)
7. Arias-Castro, E., Candes, E.J., Durand, A.: Detection of An Anomalous Cluster in A Network. The Annals of Statistics. 39, 278–304 (2011)
8. Brookmeyer, R., Johnsona, E., Ziegler-Grahamb, K., Arrighic, H.M.: Forecasting the Global Burden of Alzheimers Disease. Alzheimer's and Dementia. 3, 186–191 (2007)
9. Mintun, M.A., Larossa, G.N., Sheline, Y.I., Dence, C.S., Lee, S.Y., Mach, R.H., Klunk, W.E., Mathis, C.A., DeKosky, S.T., Morris, J.C.: [11C] PIB in a Nondemented Population Potential Antecedent Marker of Alzheimer Disease. Neurology. 67, 446–452 (2006)
10. Zhu, X., Rabbat, M.: Approximating Signals Supported on Graphs. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 3921–3924. (2012)
11. Shuman, D.I., Ricaud, B., Vandergheynst, P.: A Windowed Graph Fourier Transform. In: IEEE Statistical Signal Processing Workshop (SSP), pp. 133–136. (2012)
12. Bougleux, S., Elmoataz, A., Melkemi, M.: Discrete Regularization on Weighted Graphs for Image and Mesh Filtering. In: Scale Space and Variational Methods in Computer Vision, pp. 128–139. (2007)
13. Spielman, D.A.: Spectral Graph Theory and Its Applications. In: 48th Annual IEEE Symposium on Foundations of Computer Science, pp. 29–38. (2007)
14. Tipping, E.M., Bishop, C.M.: Probabilistic Principal Component Analysis. Jour. of the Royal Stat. Soci.: Series B (Stat. Meth.) 61, 611–622 (1999)
15. Sharpnack, J., Singh, A.: Identifying Graph-Structured Activation Patterns in Networks. In: Proc. of Neural Info. Proc. Sys. (2010)
16. Grimmett, G.: Probability on Graphs: Random Processes on Graphs and Lattices (Vol. 1). Cambridge University Press (2010)
17. Paredes, J.L., Wang, Z., Arce, G.R., Sadler, B.M.: Compressive Matched Subspace Detection. In: Proc. 17th European Signal Processing Conf., pp. 120–124. (2009)
18. Bassett, D.S., Bullmore, E.: Small-World Brain Networks. The Neuroscientist. 12, 512-523 (2006)
19. Watts, D.J., Strogatz, S.H.: Collective Dynamics of "Small-World" Networks. Nature. 2, 393–440 (1998)
20. Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., Andrews-Hanna, J.R., Sperling R.A., Johnson, K.A.: Cortical Hubs Revealed by Intrinsic Functional Connectivity: Mapping, Assessment of Stability, and Relation to Alzheimer's Disease. The Journal of Neuroscience. 29, 1860–1873 (2009)
21. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. Neuroimage. 15, 273-289 (2002)
22. Huang, S., Li, J., Sun, L., Liu, J., Wu, T., Chen, K., Fleisher, A., Reiman E., Ye, J.: Learning Brain Connectivity of Alzheimer's Disease from Neuroimaging Data. Advances in Neural Information Processing Systems. 22, 808–816 (2009)
23. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press (2000)
24. Hu, C., Cheng, L., Sepulcre, J., Fakhri, G.E., Lu, Y.M., and Li, Q.: A Graph Theoretical Regression Model for Brain Connectivity Learning of Alzheimer's Disease. To Appear In: Int. Symp. on Biomedical Imaging. (2013)