TOPOLOGICAL DATA RECOVERY UNDER COMPRESSED SENSING

Moo K. Chung[1,3,4,*]   *Hyekyoung Lee*[1,2]   *Peter T. Kim*[5]   *Jong Chul Ye*[6]

[1]Department of Brain and Cognitive Sciences
[2]Department of Nuclear Medicine, Seoul National University, Korea.
[3]Department of Biostatistics and Medical Informatics,
[4]Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin, Madison, WI, USA.
[5]Department of Mathematics and Statistics, University of Guelph, Canada
[6]Department of Bio and Brain Engineering, KAIST, Korea.

ABSTRACT

For functional measurements, the test statistic of the measurements is usually constructed at every points in space and thresholded to determine the regions of significant signals. This thresholding produces a small patch of regions around the critical values of the test statistic. It is known that the probability of the critical values bigger than a specific threshold can be computed as the expectation of the Euler characteristic of the patch. Motivated by this topological connection, we present a new computational framework of modeling various functional measurements as topological objects.

The level set associated with functional measurements can be approximated using a simplicial complex consisting of nodes and links. The existence of links basically determine the underlying topological structure of the signal. The strength of links can be modeled using an underdetermined linear model. By incorporating sparsity into the model, the links can be sparsely obtained making interpretation and visualization of the simplicial complex easier. The main contribution of this paper is showing the relationship between sparse topological structures to the sparse regression framework. We apply this novel framework in constructing a structural brain network model.

Index Terms— compressed sensing, LASSO, persistent homology, partial correlation, network

1. INTRODUCTION

In science and in particular medical imaging, it is usually assumed that the i-th functional measurement $f_i(x)$ at position $x \in \mathcal{M} \subset \mathbb{R}^d$ follows

$$f_i(x) = \mu(x) + \epsilon_i(x), \qquad (1)$$

where μ is the unknown mean signal to be estimated, ϵ_i is noise and $\mathcal{M}$ is the underlying manifold where the data is observed [1, 2, 3, 4, 5, 6]. The unknown signal is usually estimated by various spatial smoothing techniques over the manifold $\mathcal{M}$. $\mathcal{M}$ can be the 2D brain cortical surface (Figure 1) or 3D brain network graphs (Figure 4).

In a traditional statistical framework, inference on the model (1) proceeds as follows. At each fixed point x, consider the hypotheses

---

of the form

$$H_0(x) : \mu(x) = 0 \text{ vs. } H_1(x) : \mu(x) > 0. \qquad (2)$$

We construct a test statistic $T(x)$ as a function of measurements f_i at each fixed x. We still reject $H_0(x)$ if T is larger than some threshold h. Extending the inference framework (2) to all points in $\mathcal{M}$, we have

$$H_0 = \bigcap_{x \in \mathcal{M}} H_0(x) \text{ vs. } H_1 = \bigcup_{x \in \mathcal{M}} H_1(x).$$

H_0 is the intersection of every possible null hypotheses $H_0(x)$ over $\mathcal{M}$. Then we reject H_0 if $T(x)$ is larger than some threshold h. Note that H_0 is equivalent to the condition that $\sup_{x \in \mathcal{M}} T(x) > h$.

In order to compute the type-I error associated with H_0, we need to know the distribution of the supremum of the field $T(x)$, which is not straightforward. Hence a great deal of the imaging and statistical literature have been devoted to determining the distribution of $\sup_{x \in \mathcal{M}} T(x)$ [7, 8, 9, 6, 10]. Define the excursion set as

$$A_h = \{x \in \mathcal{M} : T(x) > h\}.$$

It is known that

$$P\Big(\sup_{x \in \mathcal{M}} T(x) > h \Big) \approx \mathbb{E}\chi(A_h), \qquad (3)$$

the expectation of the the Euler characteristic of the excursion set A_h [11, 12, 13]. The relationship (3) reformulates the usual statistical inference as a topological problem. Figure 2 shows how the excursion set changes over increasing threshold levels h.

The above framework is one way of detecting a signal, however it is not necessarily the best way to characterize complex multivariate functional data such as brain MRI. Instead of looking at the topological change of the excursion set of the statistic T over increasing threshold levels h, we look at the topological change of measurements f_i first. Let

$$B_{i,h} = \{x \in \mathcal{M} : f_i(x) > h\}$$

be the excursion set associated with the i-th measurement. We can then determine the topological structure of $B_{i,h}$ first, and perform statistical inference later (Figure 2). The main tool for investigating the topological change of the excursion set is *persistent homology* [14, 15, 16]. Using Morse theory, we can determine the topological deformation of the excursion sets by tabulating the occurrence of critical values [14]. This framework is general enough for dealing with a wide variety of noisy multivariate data including brain images [17, 18], networks [19, 20] and gene expression [21].
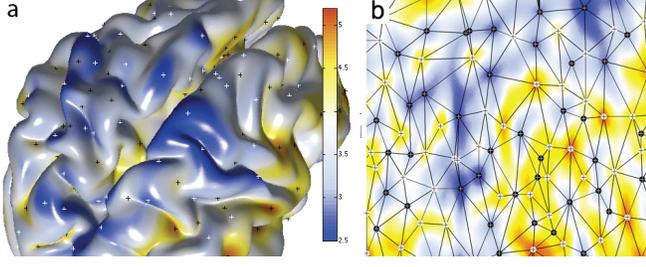
Fig. 1. (a) Heat kernel smoothing of cortical thickness in the human brain which resulted in smoothly varying estimation of cortical thickness. Critical values are then easily computed. Critical values are marked as the white (black) crosses for local minimums (maximums). (b) The simplicial complex of the underlying topology of critical values can be constructed using the Delaunay triangulation [18].

2. TOPOLOGICAL DATA RECOVERY

For continuous functional measurements f_i, its excursion set $B_{i,h}$ can be approximated using a collection of d-dimensional simplexes. Figure 1 shows an example of approximating 2D cortical thickness measures using the Delaunay triangulation [18]. On the other hand, networks are inherently topological since the underlying graph data structures are simplicial complexes [19]. So we can simply treat $B_{i,h}$ as a simplicial complex or equivalently a network consisting of nodes and links. The topological features of $B_{i,h}$ that persist over the changing threshold reflects unique characteristics of the underlying network. However, what we are really interested in is the persistent topological features that are common across different excursion sets $B_{1,h}, \cdots, B_{n,h}$. This requires a new statistical framework for network modeling. From now on, we will explicitly treat $B_{i,h}$ as a network and focus on network modeling but our new framework should work for any type of measurements.

Network modeling. Let p be the number of nodes in the network. In most applications, the number of nodes are expected to be larger than the number of observations n, which gives an underdetermined system. The i-th measurement f_i is then discretely sampled at p nodes, which we will simply index by integers. To simplify the notation, we denote $x_{ij} = f_i(j)$. At node j, we have the random variable x_j, which is realized by the random sample $x_{1j}, \cdots, x_{nj}$. We will denote this realization as $\mathbf{x}_j = (x_{1j}, \cdots, x_{nj})'$. The collection of random variables x_j are assumed to be distributed with mean zero and covariance $\Sigma = (\sigma_{ij})$ i.e.

$$\mathbb{E}x_i = 0, \ \mathbb{E}(x_i x_j) = \sigma_{ij}.$$

If $\mathbb{E}x_i \neq 0$, we can always center the data by translation. The correlation γ_{ij} between the two nodes i and j is given by

$$\gamma_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

By thresholding the correlation, we can establish a link between two nodes. However, there is a problem with this simplistic approach in that it fails to explicitly factor out the confounding effect of other nodes. To remedy this problem, partial correlations can be used in factoring out the dependency of other nodes [22, 23, 24, 25, 26].
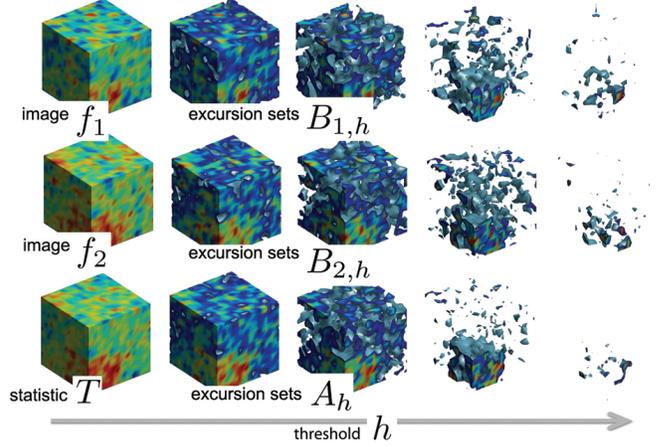
Fig. 2. (a) Given 3D functional measures $f_i = \mu + \epsilon_i$, we detect the regions of significant signal $\mu > 0$. In ~~the~~ random field theory [13], we compute the test statistic T out of f_i and determine the topological change of the excursion set $A_h = \{x \in \mathcal{M} : T(x) > h\}$ as we increase the threshold h. (b) This process determines the overall type-I error. On the other hand, what we are proposing is to determine the topological structure of the individual excursion sets $B_{i,h} = \{x \in \mathcal{M} : f_i(x) > h\}$ first. Then construct a statistical test on the topological change of $B_{i,h}$ as we change threshold h.

If we denote the inverse covariance matrix as $\Sigma^{-1} = (\sigma^{ij})$, the *partial correlation* between the nodes i and j while factoring out the effect of all other nodes is given by

$$\rho_{ij} = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}. \tag{4}$$

Equivalently, we can compute the partial correlation *via* a linear model as follows. Consider a linear model of correlating measurement at node i to all other nodes:

$$x_i = \sum_{j \neq i} \beta_{ij} x_j + \epsilon_i. \tag{5}$$

The parameters β_{ij} are estimated my minimizing the sum of squared residual of (5)

$$L(\beta) = \sum_{i=1}^{p} \|\mathbf{x}_i - \sum_{j \neq i} \beta_{ij}\mathbf{x}_j\|^2 \tag{6}$$

in a least squares fashion. If we denote the least squares estimator by $\widehat{\beta_{ij}}$, the residuals are given by

$$r_i = x_i - \sum_{j \neq i} \widehat{\beta_{ij}} x_j. \tag{7}$$

The partial correlation is then obtained by computing the correlation between the residuals of the model fit (5) [22, 27, 26]:

$$\rho_{ij} = \mathbb{E}\big[(r_i - \mathbb{E}r_i)(r_j - \mathbb{E}r_j)\big].$$

The minimization of (6) is exactly given by solving the normal equation:

$$\mathbf{x}_i = \sum_{j \neq i} \beta_{ij}\mathbf{x}_j, \tag{8}$$

which can be turned into standard linear form $y = A\beta$ [28]. Note that (8) can be written as

$$\mathbf{x}_i = \underbrace{[\mathbf{x}_1, \cdots, \mathbf{x}_{i-1}, \mathbf{0}, \mathbf{x}_{i+1}, \cdots, \mathbf{x}_p]}_{\mathbf{X}_{-i}} \underbrace{\begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{ip} \end{pmatrix}}_{\beta_i},$$

where $\mathbf{0}_{n \times 1}$ is a column vector of all zero entries. Then we have

$$\underbrace{\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{pmatrix}}_{y_{np \times 1}} = \underbrace{\begin{pmatrix} \mathbf{X}_{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{-2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{-p} \end{pmatrix}}_{A_{np \times p^2}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}}_{\beta_{p^2 \times 1}}, \quad (9)$$

where A is a block diagonal matrix and $\mathbf{0}_{n \times p}$ is a matrix of all zero entries.

Compressed sensing. However, there is a serious problem with the least squares estimation framework. Since $n \ll p$, this is a significantly underdetermined system. This is also related to the covariance matrix Σ being singular so we cannot just invert the covariance matrix in (4). So we need to regularize (9) by incorporating l_1 LASSO-penalty J [29, 26, 28]:

$$J = \sum_{i,j} |\beta_{ij}|.$$

The sparse estimation of β_{ij} is then given by minimizing $L + \lambda J$. Since there is dependency between y and A, (9) is not exactly a standard compressed sensing problem. Nevertheless, as a first order exploratory analysis, we will undergo the investigation as if they are independent as has been done in others [26, 28]. It should be intuitively understood that the sparsity makes the linear equation (8) less underdetermined. The larger the value of λ, the more sparse the underlying topological structure gets. Since

$$\rho_{ij} = \beta_{ij}\sqrt{\frac{\sigma^{ii}}{\sigma^{jj}}},$$

the sparsity of β_{ij} directly corresponds to the sparsity of ρ_{ij}, which is the strength of the link between nodes i and j [26, 28]. Once the sparse partial correlation matrix ρ is obtained, we can simply link nodes i and j, if $\rho_{ij} > 0$ and assign the weight ρ_{ij} to the edge. This way, we obtain the weighted graph. To simplify the problem, we will only consider positive partial correlations ρ^+ (Figure 3). Since the partial correlation matrix is likely to be very sparse, the resulting weighted graph will have an easily interpretable topological structure.

3. BRAIN NETWORK MODELING

The majority of functional and structural connectivity studies in brain imaging are usually performed following the standard analysis framework [30, 31, 32, 33]. From 3D whole brain images, n regions of interest (ROI) are identified and serve as the nodes of the brain network. Measurements at ROIs are then correlated in a pair-wise fashion to produce the connectivity matrix of size $n \times n$. The connectivity matrix is then thresholded to produce the adjacency matrix consisting of zeros and ones that define the link between two
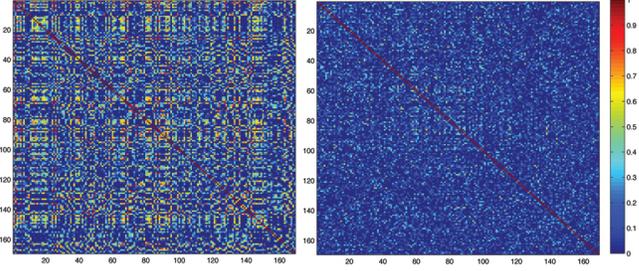
Fig. 3. Partial correlation estimation using least squares estimation (left) and LASSO with $\lambda = 100$ (right). Only positive correlations are shown. The LASSO framework reduces the number of links in the network by forcing the sparsity on partial correlation.

nodes. The binarized adjacency matrix is then used to construct the brain network. Then various graph complexity measures such as degree, clustering coefficients, entropy, path length, hub centrality and modularity are defined on the graph and the subsequent statistical inference is performed on these complexity measures.

For a large number of nodes, simple thresholding of correlation will produce a large number of links which makes the interpretation difficult. For example, for 3×10^5 voxels in an image, we can possibly have a total of 9×10^{10} links in the graph. For this reason we used the compressed sensing framework in obtaining a far smaller number of significant links.

Application. We have applied the method to a group of 33 normal control (NC) subjects. T1-weighted MRIs were collected using a 3T GE SIGNA scanner. Details on the image preprocessing pipelines are explained in [34]. The Jacobian determinant of deformation from individual MRI to a template is computed at each voxel. 169 voxels in the white matter are uniformly selected to form nodes of the network. We have sparsely estimated the partial correlation of the Jacobian determinants across nodes (Figure 3). This is a huge l_1-minimization problem and requires to estimate a total of 169^2 parameters. For instance, the matrix A is of size 5577×28561. We have used the interior-point method with various λ values in sparsely estimating ρ^+ [35]. Using ρ_{ij}^+ as link weights, we can construct the weighted graphs (Figure 4).

4. DISCUSSIONS

We have presented a novel framework that formulates signal detection as an underlying topological structure recovery. Functional measurements are represented as simplicial complexes or networks. The strength of the links between nodes is measured using partial correlations. The compressed sensing framework is used to sparsely estimate the partial correlations and obtain sparse topological structures. The sparse topological structures are easier to interpret as demonstrated in brain network modeling.

However, this new framework has a serious computational bottleneck. For n measurements over p nodes, it is required that we solve a linear system with an extremely large A matrix of size $np \times p^2$, so that the complexity of the problem increases by a factor of p^3! Consequently, for a large number of nodes, the problem immediately becomes almost intractable for a small computer. For example, for 1 million nodes, we have to compute 1 trillion possible pairwise relationships between nodes. One practical solution is to modify (5)
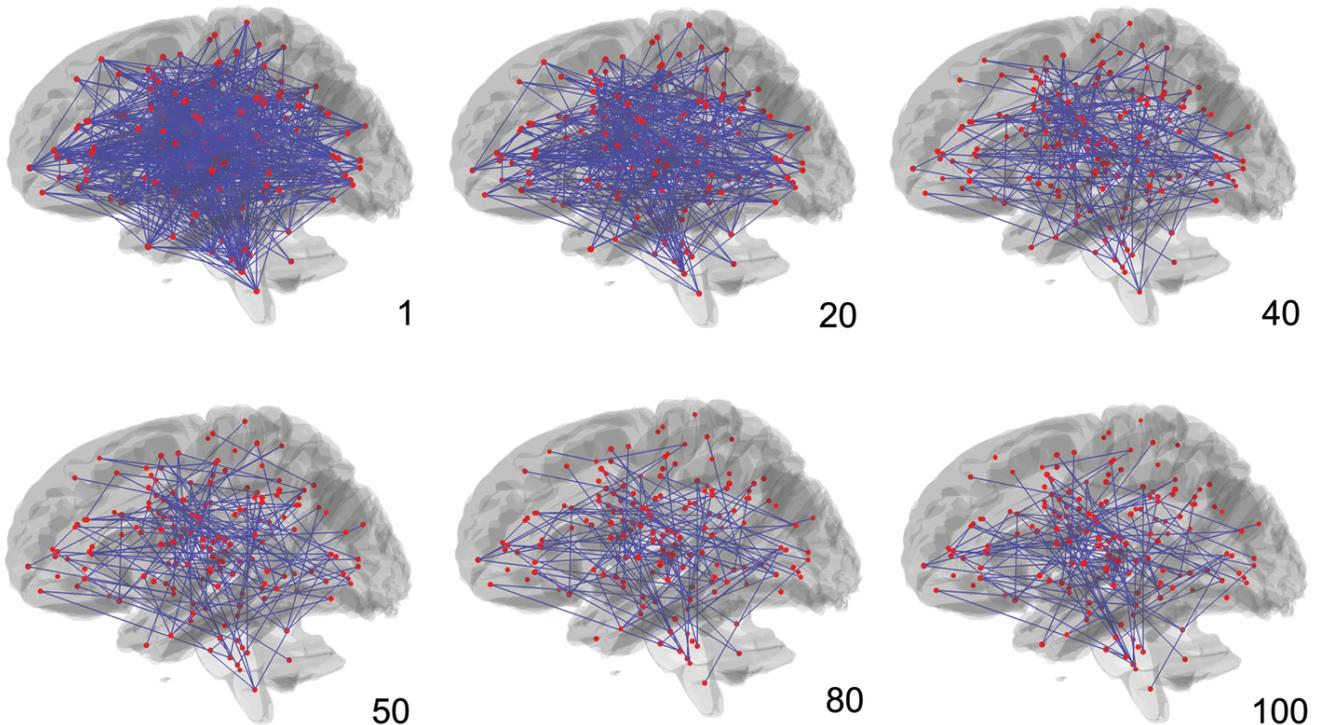
Fig. 4. Structural brain network of 33 control subjects. LASSO-penalty is used for various λ values from 1 to 100. Increasing the λ value increases the sparsity of connections and, in turn, simplifies the topological structure of the network. Persistent topological features such as Betti numbers over increasing λ can be used in the characterization of brain networks [20]

so that the measurement at node i is represented more sparsely over some possible index set S_i:

$$x_i = \sum_{S_i} \beta_{ij} x_j + \epsilon_i.$$

making the problem substantially smaller.

An alternate approach is to simply follow the *homotopy path*, which enables to add network links one by one with a very limited increase of computational complexity so we do not need to compute β repeatedly from scratch [36, 37, 38]. The trajectory of the optimal solution β in LASSO follows a piecewise linear path as we change λ. By tracing the linear path, we can substantially reduce the computational burden of reestimating β when λ changes.

5. REFERENCES

[1] M.K. Chung, K.J. Worsley, T. Paus, D.L. Cherif, C. Collins, J. Giedd, J.L. Rapoport, , and A.C. Evans, "A unified statistical approach to deformation-based morphometry," *NeuroImage*, vol. 14, pp. 595–606, 2001.

[2] K.J. Friston., "A short history of statistical parametric mapping in functional neuroimaging," Tech. Rep. Technical report, Wellcome Department of Imaging Neuroscience, ION, UCL., London, UK., 2002.

[3] S.J. Kiebel, J.-P. Poline, K.J. Friston, A.P. Holmes, and K.J. Worsley, "Robust smoothness estimation in statistical paramet-ric maps using standardized residuals from the general linear model," *NeuroImage*, vol. 10, pp. 756–766, 1999.

[4] S.C. Joshi, *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain Sub-Manifolds*, 1998.

[5] M.I. Miller, A. Banerjee, G.E. Christensen, S.C. Joshi, N. Khaneja, U. Grenander, and L. Matejic, "Statistical methods in computational anatomy," *Statistical Methods in Medical Research*, vol. 6, pp. 267–299, 1997.

[6] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans, "A unified statistical approach for determining significant signals in images of cerebral activation," *Human Brain Mapping*, vol. 4, pp. 58–73, 1996.

[7] D.Q. Naiman, "Volumes for tubular neighborhoods of spherical polyhedra and statistical inference," *Ann. Statist.*, vol. 18, pp. 685–716, 1990.

[8] T. Nichols and S. Hayasaka, "Controlling the familywise error rate in functional neuroimaging: a comparative review," *Stat Methods Med. Res.*, vol. 12, pp. 419–446, 2003.

[9] J.E. Taylor and K.J. Worsley, "Random fields of multivariate test statistics, with applications to shape analysis," *Annals of Statistics*, vol. 36, pp. 1–27, 2008.

[10] K.J. Worsley, S. Marrett, P. Neelin, and A.C. Evans, "Searching scale space for activation in pet images," *Human Brain Mapping*, vol. 4, pp. 74–90, 1996.

[11] R.J. Adler, *The Geometry of Random Fields*, John Wiley & Sons, 1981.

[12] J.E. Taylor and K.J. Worsley, "Detecting sparse signals in random fields, with an application to brain mapping," *Journal of the American Statistical Association*, vol. 102, pp. 913–928, 2007.

[13] K.J. Worsley, "Detecting activation in fmri data.," *Statistical Methods in Medical Research.*, vol. 12, pp. 401–418, 2003.

[14] H. Edelsbrunner and J. Harer, "Persistent homology - a survey," in *Surveys on discrete and computational geometry: twenty years later: AMS-IMS-SIAM Joint Summer Research Conference*, 2008, vol. 453, p. 257.

[15] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," *Discrete and Computational Geometry*, vol. 28, pp. 511–533, 2002.

[16] A.J. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete and Computational Geometry*, vol. 33, pp. 249–274, 2005.

[17] M.K. Chung, P. Bubenik, and P.T. Kim, "Persistence diagrams of cortical surface data," *Proceedings of the 21st International Conference on Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science (LNCS)*, vol. 5636, pp. 386–397, 2009.

[18] M.K. Chung, V. Singh, P.T. Kim, K.M. Dalton, and R.J. Davidson, "Topological Characterization of Signal in Brain Images Using Min-Max Diagrams," *MICCAI, Lecture Notes in Computer Science (LNCS)*, vol. 5762, pp. 158–166, 2009.

[19] D. Horak, S. Maletić, and M. Rajković, "Persistent homology of complex networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, pp. P03034, 2009.

[20] H. Lee, M.K. Chung, H. Kang, B.-N. Kim, and Lee D.S., "Discriminative persistent homology of brain networks," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2011, submitted.

[21] Dequent M-L. Mileyko Y. Edelsbrunner, H. and O. Pourquie, "Assessing periodicity in gene expression as measured by microarray data," *Preprint*, 2008.

[22] Y. He, Z.J. Chen, and A.C. Evans, "Small-world anatomical networks in the human brain revealed by cortical thickness from MRI," *Cerebral Cortex*, vol. 17, pp. 2407–2419, 2007.

[23] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional MRI," *Neuroimage*, vol. 32, pp. 228–237, 2006.

[24] S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman, and J. Ye, "Learning Brain Connectivity of Alzheimer's Disease from Neuroimaging Data," in *Proceedings of Neural Information Processing Systems Conference (NIPS)*, 2009.

[25] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, and E. Reiman, "Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation," *NeuroImage*, vol. 50, pp. 935–949, 2010.

[26] J. Peng, P. Wang, N. Zhou, and J. Zhu, "Partial correlation estimation by joint sparse regression models," *Journal of the American Statistical Association*, vol. 104, pp. 735–746, 2009.

[27] J.P. Lerch, K. Worsley, W.P. Shaw, D.K. Greenstein, R.K. Lenroot, J. Giedd, and A.C. Evans, "Mapping anatomical correlations across cerebral cortex (MACACC) using cortical thickness from MRI," *Neuroimage*, vol. 31, pp. 993–1003, 2006.

[28] H. Lee, D.S.. Lee, H. Kang, B.-N. Kim, and M.K. Chung, "Sparse brain network recovery under compressed sensing," *IEEE Transactions on Medical Imaging*, p. in press, 2011.

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, pp. 267–288, 1996.

[30] G. Gong, Y. He, L. Concha, C. Lebel, D.W. Gross, A.C. Evans, and C. Beaulieu, "Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography," *Cerebral Cortex*, vol. 19, pp. 524–536, 2009.

[31] P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V.J. Wedeen, R. Meuli, and J.P. Thiran, "Mapping human whole-brain structural networks with diffusion MRI," *PLoS One*, vol. 2, no. 7, pp. e597, 2007.

[32] A. Fornito, A. Zalesky, and E.T. Bullmore, "Network scaling effects in graph analytic studies of human resting-state fMRI data," *Frontiers in Systems Neuroscience*, vol. 4, pp. 1–16, 2010.

[33] A. Zalesky, A. Fornito, I.H. Harding, L. Cocchi, M. Y"ucel, C. Pantelis, and E.T. Bullmore, "Whole-brain anatomical networks: Does the choice of nodes matter?," *Neuroimage*, vol. 50, pp. 970–983, 2010.

[34] J.L. Hanson, M.K. Chung, B.B. Avants, E.A. Shirtcliff, J.C. Gee, R.J. Davidson, and S.D. Pollak, "Early Stress Is Associated with Alterations in the Orbitofrontal Cortex: A Tensor-Based Morphometry Investigation of Brain Structure and Behavioral Risk," *Journal of Neuroscience*, vol. 30, pp. 7466–7477, 2010.

[35] S.J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An Interior-Point Method for Large-Scale l_1-Regularized Least Squares," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, pp. 606–617, 2008.

[36] D.L. Donoho and Y. Tsaig, *Fast solution of l_1-norm minimization problems when the solution may be sparse*, Citeseer, 2006.

[37] M.D. Plumbley, "Geometry and homotopy for l_1 sparse representations," *Proceedings of SPARS*, vol. 5, pp. 206–213.

[38] M.R. Osborne, B. Presnell, and B.A. Turlach, "A new approach to variable selection in least squares problems," *IMA journal of numerical analysis*, vol. 20, pp. 389–404, 2000.