

Exploiting Hidden Persistent Structures in Multivariate Tensor-Based Morphometry and Its Application to Detecting White Matter Abnormality in Maltreated Children

Moo K. Chung¹, Jamie L. Hanson¹, Hyekyung Lee², Nagesh Adluru¹,
Andrew L. Alexander¹, Richard J. Davidson¹, Seth D. Pollak¹

¹University of Wisconsin-Madison, USA

²Seoul National University, Korea

mkchung@wisc.edu

Abstract. We present novel multivariate tensor-based morphometry (TBM) for characterizing white matter abnormalities. Traditionally TBM is used in quantifying tissue volume changes in a massive univariate fashion. At each voxel, the Jacobian determinant obtained from TBM is used as the response variable in a general linear model (GLM) and a test statistic is constructed. However, this obvious approach cannot be used in testing, for instance, if the change in one voxel is related to other voxels. To address this limitation of univariate-TBM, we propose a novel multivariate framework for more complex relational hypotheses across brain regions. To develop multivariate-TBM, it is necessary to regularize ill-conditioned covariance matrix by incorporating sparse penalty. Unfortunately, most sparse models like compressed sensing, sparse likelihood and LASSO cause a serious computational bottleneck. The computational bottleneck can be bypassed by exploiting hidden persistent structures in the sparse models. The proposed methods are applied to quantify abnormal white matter in maltreated children to show multivariate-TBM combined with persistent homology can extract additional information that cannot be obtained in univariate-TBM.

1 Introduction

Tensor-based morphometry (TBM) uses the spatial derivatives of deformation fields obtained during nonlinear image registration [3,4,27,8]. The morphological tensor maps are subsequently computed and used to quantify variations in high order morphometric changes at each voxel. From these tensor maps, statistical parametric maps are created for a group of subjects in the 3D whole brain volume, on the 2D cortical surface [26,2,20,32] or on the surface of the brain substructures such as the hippocampus and amygdala [30,31].

Previous TBM analyses have been massively univariate in that response variables are fitted using a linear model at each voxel producing massive number

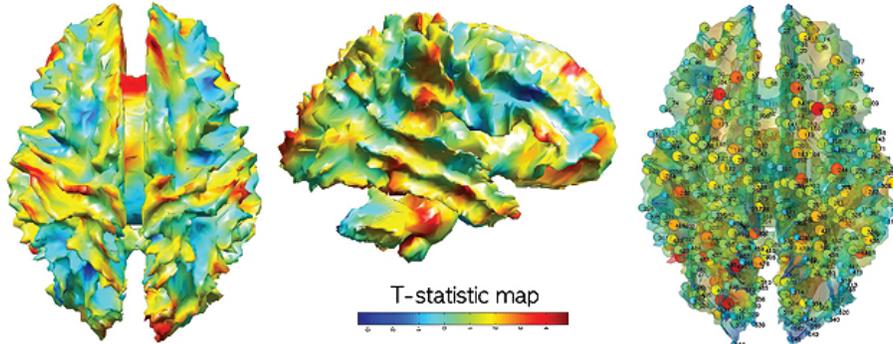


Fig. 1: Left, Middle: T -statistic map of group differences (PI-controls) on Jacobian determinants. Red regions above 4.86 are considered as statistically significant at 0.05 (corrected). Right: 548 uniformly sampled nodes where multivariate-TBM will be performed. The nodes are sparsely sampled in the template to guarantee there is no spurious high correlation due to proximity between nodes.

of test statistics (Figure 1). However, univariate-TBM is ill-suited for testing more complex hypotheses about multiple anatomical regions. For example, the univariate-TBM cannot answer how the volume increase in one voxel is related to other voxels. To address this type of more complex relational hypothesis across different brain regions, multivariate-TBM is needed.

Motivated by the limitation of traditional univariate-TBM, we present a novel multivariate framework for testing more complex relational hypotheses for multiple brain regions. We propose to correlate the Jacobian determinant across different voxels and quantify how the volume change in one voxel is correlated to the volume changes in other voxels. However, the direct application of existing multivariate statistical methods exhibits serious defects in applying to the whole brain regions due to the *small- n large- p problem* [11,23,29]. Specifically, the number of voxels p are substantially larger than the number of subjects n so the often used maximum likelihood estimation of the covariance matrix shows the rank deficiency and it is no longer positive definite. In turn, the estimated correlation matrix is not considered as a good approximation to the true correlation matrix. The small- n large- p problem can be solved by regularizing the ill-conditioned covariance matrix by sparse regularization terms. Unfortunately, many sparse regression frameworks such as compressed sensing, sparse likelihood and LASSO (least absolute shrinkage and selection operator) cause a serious computational bottleneck when trying to apply the methods to the whole brain.

Sparse model \mathcal{A} is usually parameterized by a tuning parameter λ that controls the sparsity of the representation. So the sparse model $\mathcal{A}(\lambda)$ can be viewed as a function of λ . Increasing the sparse parameter λ makes the representation more sparse. To overcome the computational bottleneck in obtaining sparse so-

lutions, we propose to identify the persistent homological structures in $\mathcal{A}(\lambda)$ for reducing computational complexity. Within the persistent homology framework, $\mathcal{A}(\lambda)$ is *persistent* if it has the nested subset structure under changes in λ value [19,9,13]. Then by exploiting the hidden persistent homology in $\mathcal{A}(\lambda)$, we will show that it is possible to completely bypass the computational bottlenecks and speed up the computation by the factor of more than ten-thousand times.

The proposed framework is applied in characterizing abnormal white matter alterations in children who experienced maltreatment while living in post-institutional (PI) settings in Eastern Europe and China before being adopted by families in the US. These children will be compared to age-matched children who did not experience maltreatment.

The main contributions of the paper are (1) the introduction of new multivariate-TBM *via* sparse regression, (2) the identification of hidden persistent homological structures in sparse regression such as LASSO and sparse likelihood (3) the utilization of persistent structures in reducing computational complexity, and (4) its novel application to clinical brain imaging data.

2 Why Sparse Models Are Needed for Multivariate-TBM?

Let $\mathbf{J}_{n \times p} = (J_{ij})$ be the data matrix of Jacobian determinant for subject i at voxel position j . The subscripts denote the dimension of matrices. Assume there are p voxels of interest and n subjects. Through the paper, we will use the following notations. The Jacobian determinants of all subjects at the j -th voxel is denoted as $\mathbf{x}_j = (J_{1j}, \dots, J_{nj})'$. The Jacobian determinants of all voxels for the i -th subject is denoted as $\mathbf{y}_i = (J_{i1}, \dots, J_{ip})'$. \mathbf{x}_j is the j -th column and \mathbf{y}_i is the i -th row of the data matrix \mathbf{J} .

If we are interested in quantifying the relationship among Jacobian determinants in every voxel simultaneously, the standard procedure is to set up a multivariate general linear model (MGLM). MGLM generalizes widely used univariate general linear models (GLM) by incorporating vector valued response and explanatory variables [1,11,33,34,24]. MGLM assumes \mathbf{y}_i are independent and identically distributed multivariate normal with mean vector μ and covariance Σ . When $p = 1$, MGLM collapses to GLM and the resulting test statistics becomes Hotelling's T^2 statistic often used for inference on vector data [25,18,12,8]. Note that the covariance matrix of \mathbf{y}_i is given by

$$\text{Cov}(\mathbf{y}_i) = \Sigma_{p \times p} = (\sigma_{kl}).$$

For a notational convenience, suppose we center the Jacobian determinant such that

$$\mathbf{y}_i \leftarrow \mathbf{y}_i - \mathbb{E}\mathbf{y}_i.$$

Basically we are subtracting the group mean from individual Jacobian maps to make $\mathbb{E}\mathbf{y}_i = 0$. Then neglecting constant terms, the log-likelihood function L of the data matrix is given by

$$L(\Sigma) = \log \det \Sigma^{-1} - \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i' \Sigma^{-1} \mathbf{y}_i \quad (1)$$

The maximum likelihood estimate (MLE) of Σ is trivially given as

$$S = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' = \frac{1}{n} \mathbf{J}'_{p \times n} \mathbf{J}_{n \times p}, \quad (2)$$

which is the sample covariance. However, there is a serious defect with MLE (2), namely the estimated covariance matrix S is ill-conditioned for $n < p$, which is true for almost all neuroimaging studies. Note that there are more voxels (p) than the number of subject (n) in most studies. Hence, MLE does not yield a good estimation in estimating the covariance matrix [11,23]. This is the main reason why MGLM was rarely employed over the whole brain and instead massive univariate approaches are still used in most neuroanatomical studies.

To remedy this small- n and large- p problem, we propose to regularize the likelihood term with L_1 -penalty and maximize the sparse likelihood:

$$L(\Sigma) = \log \det \Sigma^{-1} - \text{tr}(\Sigma^{-1}S) - \lambda \|\Sigma^{-1}\|, \quad (3)$$

where $\|\cdot\|$ is the sum of the absolute values of the elements. The tuning parameter $\lambda > 0$ controls the sparsity of the offdiagonal elements of the covariance matrix. Then we maximize L over the space of all possible symmetric positive definite matrices. (3) is a convex problem and we solve it using the graphical-lasso (GLASSO) algorithm [6,10,16]. By increasing λ , the estimated covariance matrix becomes more sparse.

3 Hidden Persistent Structures for Sparse Regressions

Since the different choice of sparse parameter λ will produce different results, we propose to use the collection of $\Sigma(\lambda)$ for every possible value of λ for the subsequent statistical inference. This avoids the problem of identifying the optimal sparse parameter that may not be optimal in practice. Unfortunately, GLASSO is fairly time consuming algorithm [10,16]. For instance, solving GLASSO for 548 nodes takes about 6 minutes on a desktop computer. To reduce the computational burden, we propose to identify hidden persistent homological structures in sparse regression and exploits their features for reducing the computational complexity [13,19].

Sparse representation \mathcal{A} is usually parameterized by a tuning parameter λ which controls the sparsity of the representation. So the representation $\mathcal{A}(\lambda)$ can be viewed as a function of λ . Since $\mathcal{A}(\lambda)$ gets more sparse as λ increases, it might be possible to construct a nested subset structure called *filtration* on the tuning parameter such that

$$\mathcal{A}(\lambda_1) \supset \mathcal{A}(\lambda_2) \supset \mathcal{A}(\lambda_3) \supset \dots \quad (4)$$

for $\lambda_1 \leq \lambda_2 \leq \dots$. In this section, we will explicitly construct such persistent structures for the first time.

3.1 Persistent Structures for Sparse Correlations

To simplify the argument, we assume the measurement vector \mathbf{x}_j at the j -th node is centered with zero mean and unit variance. These condition is achieved by centering and normalizing data such that $\mathbf{x}_i' \mathbf{x}_i = 1$ and $\sum_{i=1}^n x_{ij} = 0$. Let $\mathbf{\Gamma} = (\gamma_{jk})$ be the correlation matrix, where γ_{jk} is the correlation between the nodes j and k . The sample correlation $\hat{\gamma}_{jk} = \mathbf{x}_j' \mathbf{x}_k$ is shown to satisfy

$$\hat{\gamma}_{jk} = \arg \min_{\gamma_{jk}} \sum_{j=1}^p \sum_{k \neq j} \|\mathbf{x}_j - \gamma_{jk} \mathbf{x}_k\|_2^2. \quad (5)$$

The sparse version of (5) is the minimization of

$$F(\gamma_{jk}) = \frac{1}{2} \sum_{j=1}^p \sum_{k \neq j} \|\mathbf{x}_j - \gamma_{jk} \mathbf{x}_k\|_2^2 + \lambda \sum_{j,k=1}^p |\gamma_{jk}|. \quad (6)$$

This is the compressed sensing or LASSO type of sparse regression. By increasing $\lambda \geq 0$, the estimated correlation matrix $\hat{\mathbf{\Gamma}}(\lambda)$ becomes more sparse. The minimum of F is then achieved when

$$0 = \frac{\partial F}{\partial \gamma_{jk}} = \gamma_{jk} - \mathbf{x}_j' \mathbf{x}_k \pm \lambda.$$

The sign of λ depends on the sign of γ_{jk} . Due to this simple expression, there is no need to optimize (6) numerically using the coordinate descent learning or the active-set algorithm often used in compressed sensing [22,10]. Then for $\lambda \geq 0$, the sparse correlation estimation is given by

$$\hat{\gamma}_{jk}(\lambda) = \begin{cases} \mathbf{x}_j' \mathbf{x}_k - \lambda & \text{if } \mathbf{x}_j' \mathbf{x}_k > \lambda \\ \mathbf{x}_j' \mathbf{x}_k + \lambda & \text{if } \mathbf{x}_j' \mathbf{x}_k < -\lambda \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Using the sparse solution (7), we implicitly construct a persistent homological structure. We will basically build a graph \mathcal{G} using spare correlations. Let $A(\lambda) = (a_{ij})$ be the adjacency matrix defined as

$$a_{jk}(\lambda) = \begin{cases} 1 & \text{if } \hat{\gamma}_{jk} \neq 0; \\ 0 & \text{otherwise.} \end{cases}$$

This is equivalent to the adjacency matrix $B = (b_{jk})$ defined as

$$b_{jk}(\lambda) = \begin{cases} 1 & \text{if } |\mathbf{x}_j' \mathbf{x}_k| > \lambda; \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

The adjacency matrix B is simply obtained by thresholding the sample correlations. Then the adjacency matrices A and B induce a identical graph $\mathcal{G}(\lambda)$

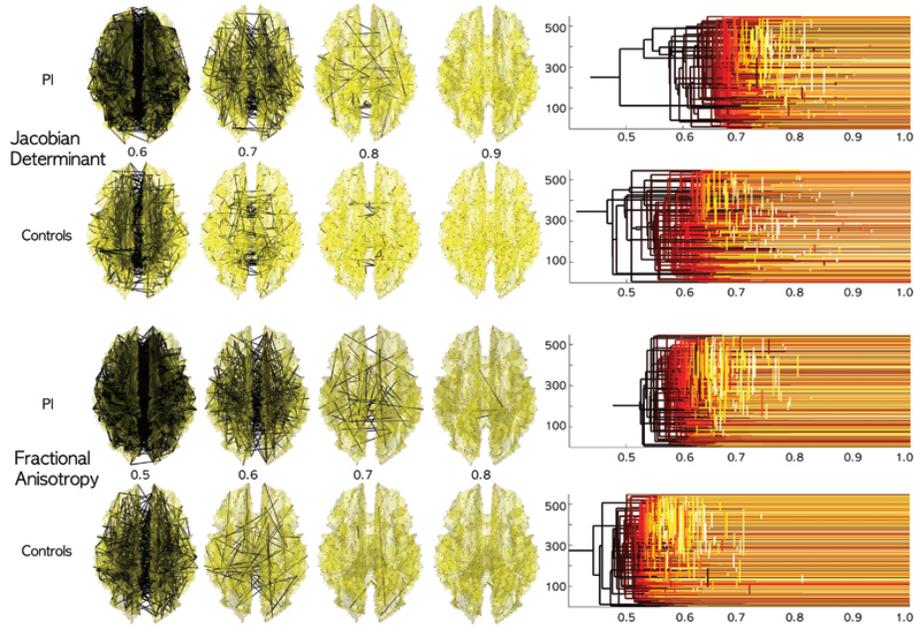


Fig. 2: Graph $\mathcal{G}(\lambda)$ obtained from sparse correlation for the Jacobian determinant from MRI and fractional anisotropy from DTI at different λ values. $\mathcal{G}(\lambda)$ forms a filtration over increasing λ . PI shows more dense network at a given filtration value. Since PI is more homogenous in the white matter region, there are more dense high correlations between nodes. The filtration over the correlation can be also visualized using a dendrogram [7,19], which also shows more dense connections for PI.

consisting of $\kappa(\lambda)$ number of partitioned subgraphs

$$\mathcal{G}(\lambda) = \bigcup_{l=1}^{\kappa(\lambda)} G_l(\lambda) \quad \text{with } G_l = \{V_l(\lambda), E_l(\lambda)\},$$

where V_l and E_l are node and edge sets respectively. Note

$$G_l \cap G_m = \emptyset \quad \text{for any } l \neq m.$$

and no two nodes between the different partitions are connected. The node and edge sets are denoted as $\mathcal{V}(\lambda) = \bigcup_{l=1}^{\kappa} V_l$ and $\mathcal{E}(\lambda) = \bigcup_{l=1}^{\kappa} E_l$ respectively. Then we have the following theorem:

Theorem 1. *The induced graph from the sparse correlation form a filtration:*

$$\mathcal{G}(\lambda_1) \supset \mathcal{G}(\lambda_2) \supset \mathcal{G}(\lambda_3) \supset \dots \quad (9)$$

for $\lambda_1 \leq \lambda_2 \leq \lambda_3$. Equivalently, the node and edge sets also form filtrations as well:

$$\mathcal{V}(\lambda_1) \supset \mathcal{V}(\lambda_2) \supset \mathcal{V}(\lambda_3) \supset \cdots, \quad \mathcal{E}(\lambda_1) \supset \mathcal{E}(\lambda_2) \supset \mathcal{E}(\lambda_3). \quad (10)$$

The proof can be easily obtained from the definition of adjacency matrix (8).

Hence we have the persistent homological structure induced from the compressed sensing type of the form (6). Figure 2 shows filtrations obtained from sparse correlations between Jacobian determinants on preselected 548 nodes in the two groups showing group difference. The sequential network construction (9) can be done in $O(p^2)$ by simply computing pairwise correlations and thresholding the sorted sample correlations sequentially.

3.2 Persistent Structures for Sparse Likelihood

The identification of a persistent homological structure out of the inverse covariance $\widehat{\Sigma}^{-1}(\lambda)$ for sparse-likelihood (3) is similar. Let $A(\lambda) = (a_{ij})$ be the adjacency matrix given by

$$a_{ij}(\lambda) = \begin{cases} 1 & \text{if } \widehat{\sigma}^{ij} \neq 0; \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

The adjacency matrix A induces a graph $\mathcal{G}(\lambda)$ consisting of $\kappa(\lambda)$ number of partitioned subgraphs

$$\mathcal{G}(\lambda) = \bigcup_{l=1}^{\kappa(\lambda)} G_l(\lambda) \quad \text{with } G_l = \{V_l(\lambda), A_l(\lambda)\}.$$

Motivated by the sparse correlation (8), let us similarly define a corresponding adjacency matrix $B(\lambda) = (b_{ij})$ as

$$b_{ij}(\lambda) = \begin{cases} 1 & \text{if } |\widehat{s}_{ij}| > \lambda; \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

The adjacency matrix B similarly induces a graph with $\tau(\lambda)$ disjoint subgraphs:

$$\mathcal{H}(\lambda) = \bigcup_{l=1}^{\tau(\lambda)} H_l(\lambda) \quad \text{with } H_l = \{W_l(\lambda), B_l(\lambda)\}.$$

Unlike the sparse correlation case, $\mathcal{G} \neq \mathcal{H}$ and we do not have full persistency on both the node and edge sets. However, the partitioned graphs are shown to be partially nested in a sense that the node sets exhibits persistency.

Theorem 2. *For any $\lambda > 0$, the adjacency matrices (11) and (12) induce the identical vertex partition so that $\kappa(\lambda) = \tau(\lambda)$ and $V_l(\lambda) = W_l(\lambda)$. Further, the node sets V_l and W_l form a filtration over the sparse parameter:*

$$V_l(\lambda_1) \supset V_l(\lambda_2) \supset V_l(\lambda_3) \supset \cdots \quad (13)$$

$$W_l(\lambda_1) \supset W_l(\lambda_2) \supset W_l(\lambda_3) \supset \cdots \quad (14)$$

for $\lambda_1 \leq \lambda_2 \leq \lambda_3$.

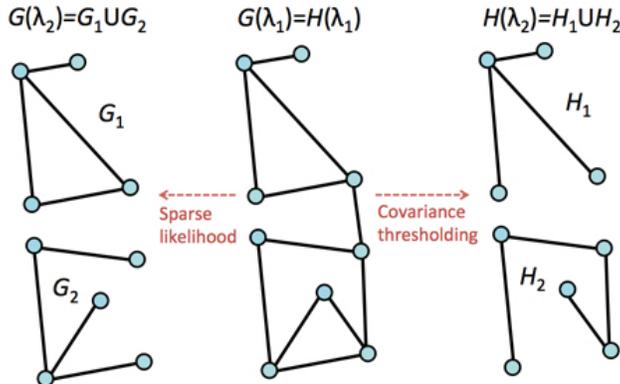


Fig. 3: Schematic of graph filtration obtained by sparse-likelihood and covariance thresholding. For $\lambda_1 \leq \lambda_2$, we have $G(\lambda_2) \subset G(\lambda_1) = H(\lambda_1) \supset H(\lambda_2)$. It can be shown that the node sets of the partitioned subnetworks between the two filtrations exactly match although the edge sets may not match (Theorem 2). Exploiting this hidden topological structure, we can drastically speed up network construction and topological computation.

From (12), it is trivial to see the filtration holds for W_l . The filtration for V_l is proved in [16]. The equivalence of the node sets $V_l = W_l$ is proved in [21]. Note that the edge sets may not form a filtration. The construction of the filtration on the node sets V_l (13) is very time consuming since we have to solve the sequence of GLASSO. For instance, for 548 node sets and 547 different filtration values, the whole filtration takes more than 54 hours in a desktop. However, using Theorem 2, we can construct the equivalent filtration on W_l by simply thresholding the sample covariance in $O(p^2)$ without the computational bottleneck encountered in GLASSO. Theorem 2 is illustrated in Figure 3 with two levels of filtration.

4 Application to Maltreated Children Study

4.1 MRI Data and Univariate-TBM

T1-weighted MRI were collected using a 3T GE SIGNA scanner for 23 children who experienced maltreatment while living in post-institutional (PI) settings in Eastern Europe and China before being adopted by families in the US, and age-matched 31 normal control subjects. The average age for PI is 11.26 ± 1.71 years while that of controls is 11.58 ± 1.61 years. There are 10 boys and 13 girls in PI, and 18 boys and 13 girls in the control subjects.

A study specific template was constructed using the diffeomorphic shape and intensity averaging technique through Advanced Normalization Tools (ANTs) [5]. Image normalization of each individual image to the template was done using symmetric normalization with cross-correlation as the similarity metric. The

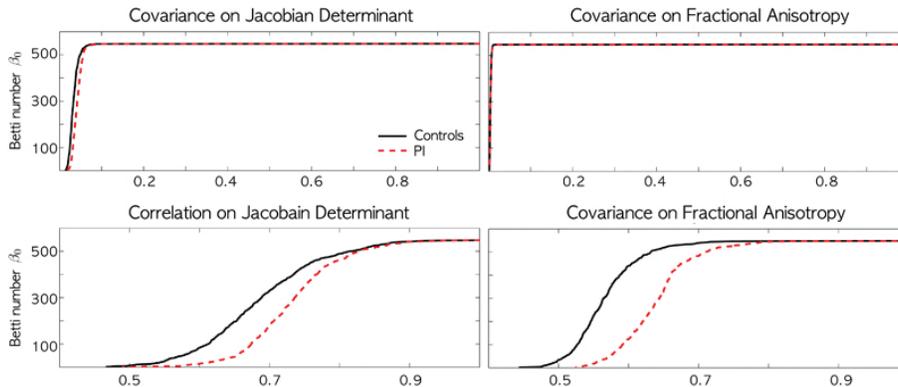


Fig. 4: The barcodes on the sparse covariance (top) and correlation (bottom) for Jacobian determinant (left) and FA (right). Unlike covariance, the correlation shows huge group separation between normal controls and post-institutionalized (PI) children (p -value < 0.001).

deformation fields are then smoothed out using Gaussian kernel with bandwidth $\sigma = 4\text{mm}$, which is equivalent to the full width at half maximum (FWHM) of 4mm . Then the Jacobian determinant of the inverse deformation was computed at each voxel.

The computed Jacobian maps were feed into univariate-GLM at each voxel for testing the group effect while accounting for nuisance covariates such as age and gender. Figure 1 shows the significant group difference between PI and controls. Any region above 4.86 or below -4.86 is considered significant at 0.05 (corrected) [34]. However, what the univariate-TBM can not test is the dependency of Jacobian determinants at two different positions. It is possible that structural abnormality at one region of the brain might be related to the other regions due to interregional dependency. For this type of more complex hypothesis, we need the proposed multivariate approach.

4.2 Multivariate-TBM via Barcodes

Since Jacobian determinants at neighboring voxels are highly correlated, we uniformly subsampled 548 nodes along the white matter boundary in order not to have spurious high correlation between two adjacent nodes (Figure 1). The number of predefined regions is still larger than most region of interest (ROI) approaches in MRI and DTI [35]. Subsequently we applied the proposed multivariate framework and obtained sparse correlations and covariance, and constructed the filtrations on them. Without solving the optimization problem (8), the sample correlation and covariance are simply thresholded to obtain the filtrations. For 547 levels of filtration, the sequence of GLASSO would take more than 54 hours in a desktop (6min. per GLASSO). But it took less than 1 min using the simple thresholding method to construct the filtrations.

A filtration is usually quantified by the barcode which plots the change of Betti numbers [9,13,19]. The first Betti number $\beta_0(\lambda)$ counts the number of connected components at the filtration value λ . Given barcode $\beta_0^i(\lambda)$ for group i , we are interested testing the null hypothesis

$$H_0 : \beta_0^1(\lambda) = \beta_0^2(\lambda) \text{ for all } \lambda \in [0, 1]$$

against the alternative hypothesis

$$H_1 : \beta_0^1(\lambda) \neq \beta_0^2(\lambda) \text{ for some } \lambda \in [0, 1].$$

Since barcodes are similar to the shape of cumulative probability distribution functions, Kolmogorov-Smirnov (KS) like test statistic can be used:

$$T = \sup_{\lambda \in [0,1]} |\beta_0^1(\lambda) - \beta_0^2(\lambda)|.$$

Since each group produces one barcode, we used the Jackknife resampling technique for inference. For a group with n subjects, one subject is removed and the remaining $n - 1$ subjects are used in computing the sparse covariance and correlations. This process is repeated for each subject to produce n covariance and correlations. Then the filtration is performed on jackknife resampled covariance and correlations (Figure 4). The Jackknife resampling produces 23 and 31 barcodes respectively for PI and controls. Then KS-like test statistic K is constructed between 23×31 pairs of barcodes. Under the null, K is expected to be zero. One-sample student T -test is then subsequently performed to show almost perfect group separation (p -value < 0.001).

5 Discussion: Connection to DTI Study

Severe stress and maltreatment during the early development is found to be related to structural abnormality in various brain regions [28,17,15,14]. Thus we expect white matter differences in not only in the Jacobian determinants but the fractional anisotropy (FA) values in DTI as well. The MRI data in this study has the corresponding DTI. The DTI acquisition are done in the same 3T GE SIGNA scanner and the acquisition parameters can be found in [blinded]. We applied the proposed methods in obtaining the sparse correlation and covariance maps in the same 548 nodes. The resulting filtration patterns also show similar pattern of rapid increase in disconnected components (Figure 2 and 4). The Jackknife-based one-sample T -test also shows significant group difference (p -value < 0.001). This results are due to consistent abnormality observed in both MRI and DTI modalities. The PI group exhibited stronger white matter homogeneity and less spatial variability compared to normal controls in both MRI and DTI measurements.

Acknowledgements

This work was supported by National Institutes of Health Research Grants MH61285 and MH68858 to S.D.P., funded by the National Institute of Mental Health and the Children's Bureau of the Administration on Children, Youth and Families as part of the Child Neglect Research Consortium, National Institute of Mental Health Grant MH84051 to R.J.D. as well as US National Institute of Drug Abuse (Fellowship DA028087) to J.L.H. The authors like to thank Matthew Arnold of University of Bristol for the discussion on the KS-test procedure and pointing out the reference [21] that has been critical for this paper. He also pointed out possible sample size dependency on the barcode constructed out of sparse correlation but that has been left for future study. Discussion with Seung-Goo Kim of Max Planck Institute is also acknowledged.

References

1. T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley., 2nd. edition, 1984.
2. A. Andrade, F Kherif, J. Mangin, K.J. Worsley, A. Paradis, O. Simon, S. Dehaene, D. Le Bihan, and J-B. Poline. Detection of fmri activation using cortical surface mapping. *Human Brain Mapping*, 12:79–93, 2001.
3. J. Ashburner and K. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.
4. J Ashburner, C. Good, and K.J. Friston. Tensor based morphometry. *NeuroImage*, 11S:465, 2000.
5. B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12:26–41, 2008.
6. O. Banerjee, L.E. Ghaoui, A. d'Aspremont, and G. Natsoulis. Convex optimization techniques for fitting sparse Gaussian graphical models. In *Proceedings of the 23rd international conference on Machine learning*, page 96, 2006.
7. G. Carlsson and F. Memoli. Persistent clustering and a theorem of J. Kleinberg. *arXiv preprint arXiv:0808.2241*, 2008.
8. 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*, 14:595–606, 2001.
9. H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.
10. J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432, 2008.
11. K.J. Friston, A.P. Holmes, K.J. Worsley, J.-P. Poline, C.D. Frith, and R.S.J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
12. C. Gaser, H.-P. Volz, S. Kiebel, S. Riehemann, and H. Sauer. Detecting structural changes in whole brain based on nonlinear deformationsapplication to schizophrenia research. *NeuroImage*, 10:107–113, 1999.
13. R. Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, 2008.

14. J.L. Hanson, M.K. Chung, B.B. Avants, K.D. Rudolph, E.A. Shirtcliff, J.C. Gee, R.J. Davidson, and S.D. Pollak. Structural variations in prefrontal cortex mediate the relationship between early childhood stress and spatial working memory. *The Journal of Neuroscience*, 32:7917–7925, 2012.
15. 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*, 30:7466–7477, 2010.
16. 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*, 50:935–949, 2010.
17. A.P. Jackowski, C. de Araujo, A. de Lacerda, de Jesus, and J. Kaufman. Neurostructural imaging findings in children with post-traumatic stress disorder: Brief review. *Psychiatry and Clinical Neurosciences*, 63:1–8, 2009.
18. S.C. Joshi. *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain Sub-Manifolds*. Ph.D. thesis. Washington University, St. Louis, 1998.
19. H. Lee, M.K. Chung, H. Kang, B.-N. Kim, and D.S. Lee. Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. *MICCAI, Lecture Notes in Computer Science*, 6892:302–309, 2011.
20. N. Lepore, C.A. Brun, M.C. Chiang, Y.Y. Chou, R.A. Dutton, K.M. Hayashi, O.L. Lopez, H.J. Aizenstein, A.W. Toga, J.T. Becker, and P.M. Thompson. Multivariate statistics of the jacobian matrices in tensor based morphometry and their application to HIV/AIDS. *Lecture Notes in Computer Science*, pages 191–198, 2006.
21. R. Mazumder and T. Hastie. Exact covariance thresholding into connected components for large-scale graphical lasso. *The Journal of Machine Learning Research*, 13:781–794, 2012.
22. J. Peng, P. Wang, N. Zhou, and J. Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104:735–746, 2009.
23. J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4:32, 2005.
24. J.E. Taylor and K.J. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *Annals of Statistics*, 36:1–27, 2008.
25. P. M. Thompson, D. MacDonald, M. S. Mega, C. J. Holmes, A. C. Evans, and A. W Toga. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *Journal of Computer Assisted Tomography*, 21:567–581, 1997.
26. P.M. Thompson, T.D. Cannon, K.L. Narr, T. van Erp, V.P. Poutanen, M. Hut-
tunen, J. Lonnqvist, C.G. Standertskjold-Nordenstam, J. Kaprio, M. Khaledy,
et al. Genetic influences on brain structure. *Nature Neuroscience*, 4:1253–1258,
2001.
27. P.M. Thompson, J.N. Giedd, R.P. Woods, D. MacDonald, A.C. Evans, and A.W
Toga. Growth patterns in the developing human brain detected using continuum-
mechanical tensor mapping. *Nature*, 404:190–193, 2000.
28. L.A. Tupler and M.D. De Bellis. Segmented hippocampal volume in children and
adolescents with posttraumatic stress disorder. *Biological psychiatry*, 59:523–529,
2006.

29. P.A. Valdés-Sosa, J.M. Sánchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360:969–981, 2005.
30. L. Wang, J.S. Swank, I. E. Glick, M. H. Gado, M. I. Miller, J. C. Morris, and J.G. Csernansky. Changes in hippocampal volume and shape across time distinguish dementia of the alzheimer type from healthy aging. *NeuroImage*, 20:667–682, 2003.
31. Y. Wang, Y. Song, P. Rajagopalan, T. An, K. Liu, Y.Y. Chou, B. Gutman, A.W. Toga, P.M. Thompson, and The Alzheimer’s Disease Neuroimaging Initiative. Surface-based tbm boosts power to detect disease effects on the brain: An n= 804 adni study. *NeuroImage*, 56:1993–2010, 2011.
32. Y. Wang, J. Zhang, B. Gutman, T.F. Chan, J.T. Becker, H.J. Aizenstein, O.L. Lopez, R.J. Tamburo, A.W. Toga, and P.M. Thompson. Multivariate tensor-based morphometry on surfaces: Application to mapping ventricular abnormalities in HIV/AIDS. *Neuroimage*, 49:2141–2157, 2010.
33. 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*, 4:58–73, 1996.
34. K.J. Worsley, J.E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *NeuroImage*, 23:S189–195, 2004.
35. A. Zalesky, A. Fornito, I.H. Harding, L. Cocchi, M. Yucel, C. Pantelis, and E.T. Bullmore. Whole-brain anatomical networks: Does the choice of nodes matter? *NeuroImage*, 50:970–983, 2010.