Presented in NIPS 2012 Workshop on Algebraic Topology and Machine Learning

Persistent Homological Structures in Compressed Sensing and Sparse Likelihood

Moo K. Chung* **Hyekyung Lee†** **Matthew Arnold‡**

*University of Wisconsin-Madison, US
†Seoul National University, Korea, ‡University of Bristol, UK

mkchung@wisc.edu hklee.brain@gmail.com matthew.arnold@bristol.ac.uk

Abstract

In this paper, we explore hidden persistent homological structures in sparse regressions. Sparse regressions are usually parameterized by tuning parameters that determine the sparsity of solutions. By treating the tuning parameters as additional dimension, we can have multi-scale representations. We can show there exist hidden persistent homological structures in these dimensions. By exploiting the hidden topological structures further, it is possible to completely bypass the computational bottlenecks that usually occur in solving the sparse regressions.

1 Introduction

A persistent data structure is one in which the state of the structure is preserved when it is modified [1]. Within the persistent homology framework, it usually refers to having the nested subset data structure under filtration [2, 3].

This paper is motivated by a simple question of if sparse data recovery techniques such as compressed sensing and sparse likelihood methods can possibly have such persistent homological data structures [4]. The answer is a resounding *yes* for two sparse representation techniques: sparse correlations and sparse likelihood.

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

$$\mathcal{A}(\lambda_1) \supset \mathcal{A}(\lambda_2) \supset \mathcal{A}(\lambda_3) \supset \cdots \tag{1}$$

for $\lambda_1 \leq \lambda_2 \leq \cdots$. In this paper, we investigate this particular persistency which is often studied in persistent homology and topological data analysis [3].

2 Methods

Through the paper, we use the following notations. Let $X_{n\times p} = (x_{ij})$ be the data matrix consisting of n observations in p nodes. Its p-dimensional row vector $\mathbf{y}_i = (x_{i1}, \cdots, x_{ip})'$ is assumed to follow multivariate normal with mean 0 and covariance $\mathbf{\Sigma} = (\sigma_{kl})$. The n-dimensional column vector at each node i is denoted as $\mathbf{x}_i = (x_{1i}, \cdots, x_{ni})'$. When $n \ll p$, this becomes a significantly underdetermined system. In this situation, $\mathbf{\Sigma}$ or its inverse are usually estimated *via* compressed sensing or sparse likelihood type of L1-norm penalty methods. [4, 5].

Sparse Correlation Matrix. To simplify the argument, we will further assume $\sigma_{ij} = 1$. The condition of zero mean and unit variance is achieved by centering and normalizing data such that $\mathbf{x}_i'\mathbf{x}_i = 1$

and $\sum_{i=1}^{n} x_{ij} = 0$. Consider the correlation matrix $\mathbf{\Gamma} = (\gamma_{jk})$ which can be estimated via linear regression

$$\mathbf{x}_j = \alpha_{jk}\mathbf{x}_k + \epsilon_j. \tag{2}$$

The least squares estimation (LSE) of γ_{jk} is given by $\widehat{\gamma}_{jk} = \widehat{\alpha}_{jk} = \mathbf{x}_j'\mathbf{x}_k$, which is the sample correlations. The sparse version of (2) is the minimization of

$$F(\gamma_{jk}) = \frac{1}{2}\sum_{j=1}^{p}\sum_{k\neq j} \parallel \mathbf{x}_j - \gamma_{jk}\mathbf{x}_k \parallel_2^2 + \lambda \sum_{j,k=1}^{p} |\gamma_{jk}|. \tag{3}$$

This is the compressed sensing type of optimization problem. By increasing λ, the estimated correlation matrix $\widehat{\mathbf{\Gamma}}(\lambda)$ 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}. Then for $\lambda \geq 0$, the sparse correlation estimation is given by

$$\widehat{\gamma}_{jk}(\lambda) = \begin{cases} \mathbf{x}_j'\mathbf{x}_k - \lambda & \text{if } \mathbf{x}_j'\mathbf{x}_k > \lambda \\ \mathbf{x}_j'\mathbf{x}_k + \lambda & \text{if } \mathbf{x}_j'\mathbf{x}_k < -\lambda \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

Due to this simple expression, there is no need to optimize (3) numerically using the coordinate descent learning or the active-set algorithm often used in compressed sensing.

Using the sparse solution (4), let us show how to construct a persistent homological structure. To simplify the problem, let just assume $\lambda \geq 0$. Let $A(\lambda) = (a_{ij})$ be the adjacency matrix given by

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

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

$$b_{jk}(\lambda) = \begin{cases} 1 & \text{if } |\mathbf{x}_j'\mathbf{x}_k| > \lambda; \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

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

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

where the whole node and edge sets are partitioned into V_l and E_l respectively. No two nodes between the different partitions are connected. From (5), it is trivial to show the partitioned graphs are nested within each partition such that

$$G_l(\lambda_1) \supset G_l(\lambda_2) \supset G_l(\lambda_3) \supset \cdots$$

for $\lambda_1 \leq \lambda_2 \leq \lambda_3$. Hence we have the persistent homological structure induced from the compressed sensing type of the form (3). *Any computation involving (3) can be done without optimization in $O(p^2 \ln p)$ time by thresholding the sorted sample correlations sequentially*. A similar but more restricted persistent homological structure can be also found in the following spare-likelihood method.

Sparse Covariance Matrix. Neglecting constant terms, the log-likelihood function L of $\mathbf{y}_i$ is given by

$$L(\mathbf{\Sigma}^{-1}) \;\; = \;\; \log\det \mathbf{\Sigma}^{-1} - \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i'\mathbf{\Sigma}^{-1}\mathbf{y}_i \tag{6}$$

The maximum likelihood estimate (MLE) of $\mathbf{\Sigma}$ is trivially given by $S = \frac{1}{n}\sum_{i=}^{n} \mathbf{y}_i\mathbf{y}_i' = \frac{1}{n}X'X$, which is the sample covariance matrix.

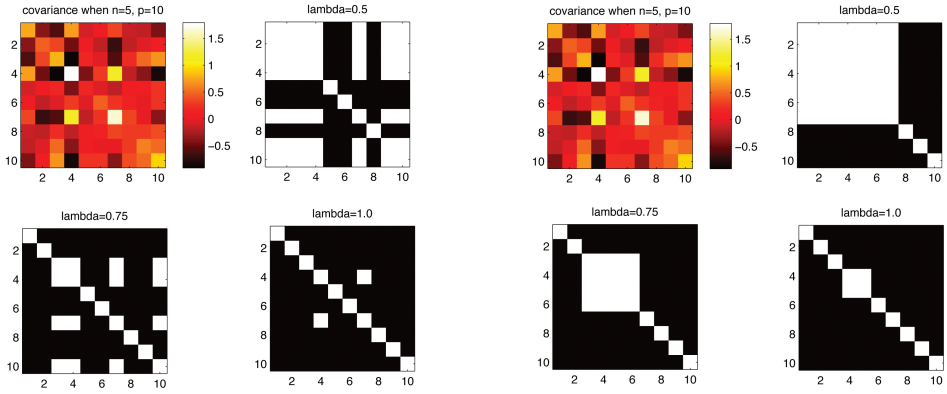

Figure 1: Left: Adjacency matrices obtained through GLASSO with increasing λ values. The persistent homological structure is self-evident. Right: A block diagonal matrix D with κ number of blocks such that $D = PAP^{-1}$ with a permutation matrix P. D also shows the persistent structure.

The sparse version of the log-likelihood is given by

$$L(\mathbf{\Sigma}^{-1}) = \log\det\mathbf{\Sigma}^{-1} - \text{tr}\Big(\mathbf{\Sigma}^{-1}S\Big) - \lambda\|\mathbf{\Sigma}^{-1}\|. \tag{7}$$

The maximization of (7) is solve by the graphical-lasso (GLASSO) algorithm [5, 6].

By increasing λ, the estimated inverse covariance matrix $\widehat{\mathbf{\Sigma}}^{-1}(\lambda) = (\widehat{\sigma}^{ij}(\lambda))$ becomes more sparse.

The construction of the persistent homological structure out of $\widehat{\mathbf{\Sigma}}^{-1}(\lambda)$ is similar to the sparse correlation case. Let $A(\lambda) = (a_{ij})$ be given by

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

Then the adjacency matrix A induces a graph $\mathcal{G}(\lambda)$ consisting of $\kappa(\lambda)$ disconnected components $\mathcal{G}(\lambda) = \bigcup_{l=1}^{\kappa(\lambda)}(V_l(\lambda), A_l(\lambda))$. Then it can be shown that the partitioned node sets are nested within each partition such that

$$V_l(\lambda_1) \supset V_l(\lambda_2) \supset V_l(\lambda_3) \supset \cdots \tag{8}$$

for $\lambda_1 \leq \lambda_2 \leq \lambda_3$ [6, 7]. The construction of the persistent structure (8) is fairly time consuming since we have to solve the sequence of GLASSO. However, it can be shown that the partitioned node sets V_l can be obtained by simply thresholding of the sample covariance $S = (s_{ij})$ [7]. Let us define a new adjacency matrix $B(\lambda) = (b_{ij})$ as

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

The adjacency matrix B similarly induces the partitioned graph $\mathcal{G}(\lambda) = \bigcup_{l=1}^{\kappa(\lambda)}(W_l(\lambda), B_l(\lambda))$. It is trivial to see

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

for $\lambda_1 \leq \lambda_2 \leq \lambda_3$. It can be further shown that $V_l = W_l$. *By the simple thresholding the sample covariance matrix, it is possible to construct the the persistent structure (8) in $O(p^2 \ln p)$ time again without the computational bottleneck associated with GLASSO*. However, in this construction, the edge sets may not exhibit nestedness.

3 Experimental Results

We illustrate our methods with two examples. In Figure 1, we randomly simulated the data matrix $X_{5\times 10}$ from the standard normal distribution. The sample covariance matrix is then feed into

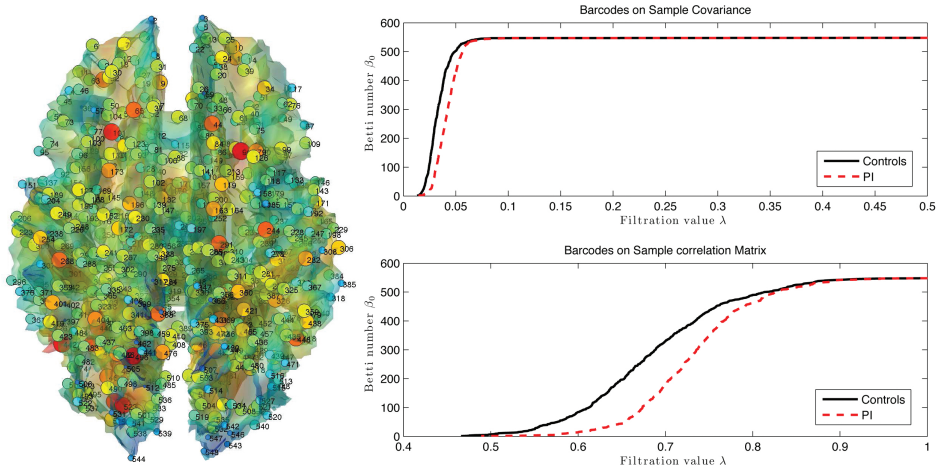

Figure 2: Jocobian determinant of deformation field are measured at 548 nodes along the human brain surface. The barcodes of the filtrations on the sample correlations and covariances show huge group separation between normal controls and post-institutionalized (PI) children.

GLASSO with different filtration values and the induced adjacency matrices show the nestedness. Theoretically only the partitioned node sets are expected to exhibit the nestedness.

Figure 2 shows the real world example based on the 3 Tesla magnetic resonance images (MRI) of 23 maltreated children who have been post-institutionalized (PI) in orphanages but later adopted to the families in US and age-matched 31 normal control subjects. The Jacobian determinants of warping individual images to the template were computed at 548 positions along the white matter brain surface [8]. The nested node sets V_l are constructed by simple thresholding on correlations and covariances. For 547 level of filtrations, the sequence of GLASSO would take more than 54 hours in a laptop (6min. per GLASSO). But it took less than 1 min using the simple thresholding method to obtain the barcodes.

References

[1] J.R. Driscoll, N. Sarnak, D.D. Sleator, and R.E. Tarjan, "Making data structures persistent," *Journal of computer and system sciences*, vol. 38, pp. 86–124, 1989.

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

[3] R. Ghrist, "Barcodes: The persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, pp. 61–75, 2008.

[4] 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*, vol. 30, pp. 1154–1165, 2011.

[5] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso.," *Biostatistics*, vol. 9, pp. 432, 2008.

[6] 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.

[7] R. Mazumder and T. Hastie, "Exact covariance thresholding into connected components for large-scale graphical lasso," *The Journal of Machine Learning Research*, vol. 13, pp. 781–794, 2012.

[8] 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*, vol. 32, pp. 7917–7925, 2012.