Latent Variable Graphical Model Selection using Harmonic Analysis: Applications to the Human Connectome Project (HCP)

Won Hwa Kim* Hyunwoo J. Kim* Nagesh Adluru Vikas Singh
http://cs.wisc.edu/~wonhwa http://cs.wisc.edu/~hwkim
University of Wisconsin-Madison

Objective

Identify conditional dependencies between random variables represented as an undirected graphical model despite unknown number of (possibly many) latent variables.

Main Idea

- Problem: Dense sample precision matrix (inverse covariance matrix), especially with high dimensional data with small sample sizes.
- Challenges: Must account for the effect from a large number of latent (unobserved) variables.
- Solution: Perform **multi-resolution analysis** of the sample precision matrix to decompose it into **dense component (low-frequency)** and **sparse component (high-frequency)**.

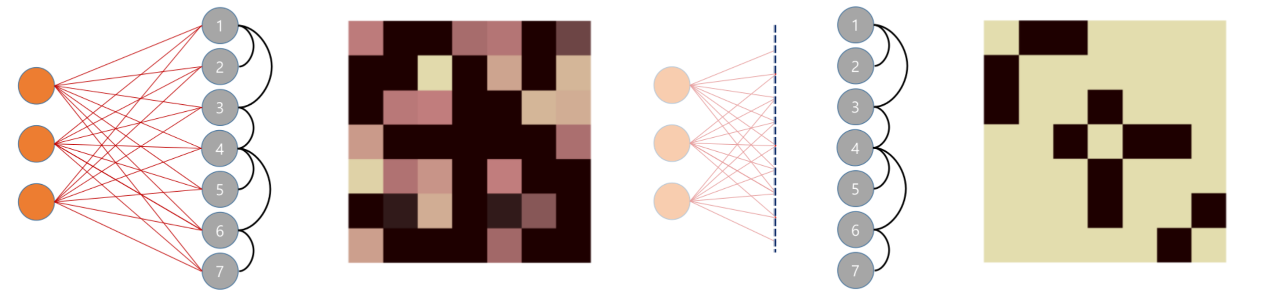


Figure : Left: effect of latent variables causing dense precision matrix, Right: removing the effect of latent varibles returning sparse precision matrix.

Preliminary: Wavelet Transform

- Transformation using wavelet bases — localized oscillating bases in both time and frequency.
- Wavelets behave as *band-pass* filters in the frequency domain.
- A mother wavelet $\psi_{s,a}$ is a function of scale s and translation a as

$$\psi_{s,a}(x) = \frac{1}{s}\psi\left(\frac{x-a}{s}\right) \tag{1}$$

- The forward wavelet transform using ψ yields wavelet coefficient $W_f(s,a)$:

$$W_f(s,a) = \langle f, \psi_s \rangle = \frac{1}{s}\int f(x)\psi^*\left(\frac{x-a}{s}\right)dx \tag{2}$$

- Inverse wavelet transform (with $C_\psi = \int \frac{|\Psi(j\omega)|^2}{|\omega|}d\omega < \infty$):

$$f(x) = \frac{1}{C_\psi}\iint W_f(s,a)\psi_{s,a}(x)\,da\,ds \tag{3}$$

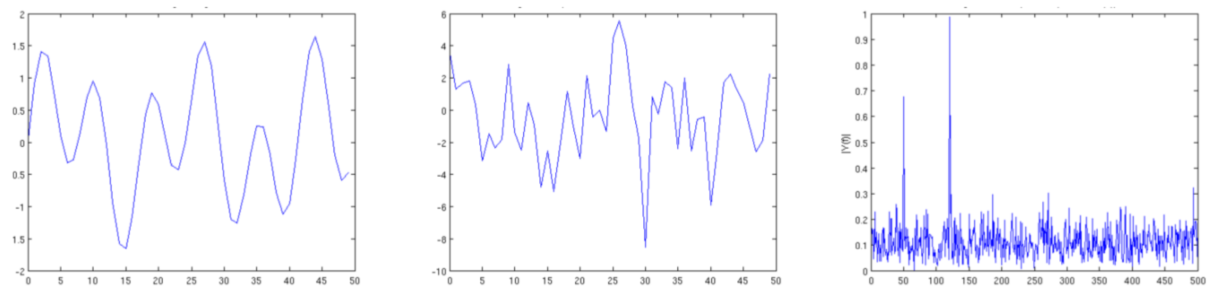


Figure : Left: true signal, Middle: signal with noise, Right: signal in the frequency space.

Construction of Wavelet in non-Euclidean Space

- Ambiguity of scale and translation of ψ in the non-Euclidean space, (represented as a graph with vertices and edges).

Construction of Wavelet in non-Euclidean Space

- Define scale in the frequency domain and translate by δ_n (Hammond 2012).
- From spectral graph theory: the eigenvector χ_l and eigenvalue λ_l pairs of graph Laplacian $\mathcal{L} = D - A$ provide an analogue of frequency domain.
- Graph Fourier Transform

$$\hat{f}(l) = \langle \chi_l, f \rangle = \sum_{n=1}^{N} \chi_l^*(n) f(n) \quad \text{and} \quad f(n) = \sum_{l=0}^{N-1} \hat{f}(l)\chi_l(n)$$

- Spectral graph wavelet is constructed by applying a band-pass filter g at various scales s and localzing it at n with a impulse function δ_n as,

$$\psi_{s,n}(m) = \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l^*(n)\chi_l(m) \tag{4}$$

- Wavelet transform on graph is defined as

$$W_f(s,n) = \langle f, \psi_s \rangle = \sum_{l=0}^{N-1} g(s\lambda_l)\hat{f}(l)\chi_l(n) \tag{5}$$

yields wavelet coefficients $W_f(s,p)$, where $\hat{f}(l) = \langle f, \chi_l \rangle$.

Harmonic Analysis of Graphical Model

- A covariance Σ / precision matrix Θ given as

$$\Sigma = V\Lambda V^T = \sum_{\ell=1}^{n} \lambda_\ell V_\ell V_\ell^T \quad \text{and} \quad \Theta = \sum_{\ell=1}^{n} \frac{1}{\lambda_\ell} V_\ell V_\ell^T = \sum_{\ell=1}^{n} \sigma_\ell V_\ell V_\ell^T$$

where $\sigma = \frac{1}{\lambda}$.
- Using the eigenvectors V, we define a basis analoguous to wavelets as

$$\psi_{\ell,s}(i,j) = g(s\sigma_\ell)V_\ell^*(i)V_\ell(j), \forall \ell \in \{1,\ldots,n\} \tag{6}$$

where $g()$ is a kernel function (band-pass filter) as in (4).
- A transform of the precision matrix using our basis yields coefficients as

$$W_{\Theta,s}(\ell) = \langle \Theta, \psi_{\ell,s} \rangle = \sigma_\ell g(s\sigma_\ell).$$

- Multi-resolution reconstruction with a non-constant weight ds/s

$$\tilde{\Theta}(i,j) = \frac{1}{C_g}\int_0^\infty \frac{1}{s}\sum_{\ell=1}^{n} W_{\Theta,s}(\ell)\psi_{\ell,s}(i,j)ds.$$

which is a multi-resolution reconstruction using different scales.

Lemma (Perfect Reconstruction of Θ)

If $\Theta \succ 0, \Theta = \Theta^T$ and kernel g satisfies the admissibility condition

$$\int_0^\infty \frac{g^2(s\sigma)}{s}ds =: C_g < \infty$$

then,

$$\frac{1}{C_g}\int_0^\infty \frac{1}{s}\sum_{\ell=1}^{n} W_{\Theta,s}(\ell)\psi_{\ell,s}(i,j)ds = \Theta(i,j)$$

Estimating the Optimal Scale for $\tilde{\Theta}$

- The reconstruction of Θ at scale s is given as

$$\tilde{\Theta} = \sum_{\ell=1}^{n} \sigma_\ell g^2(s\sigma_\ell)V_\ell V_\ell^T \tag{7}$$

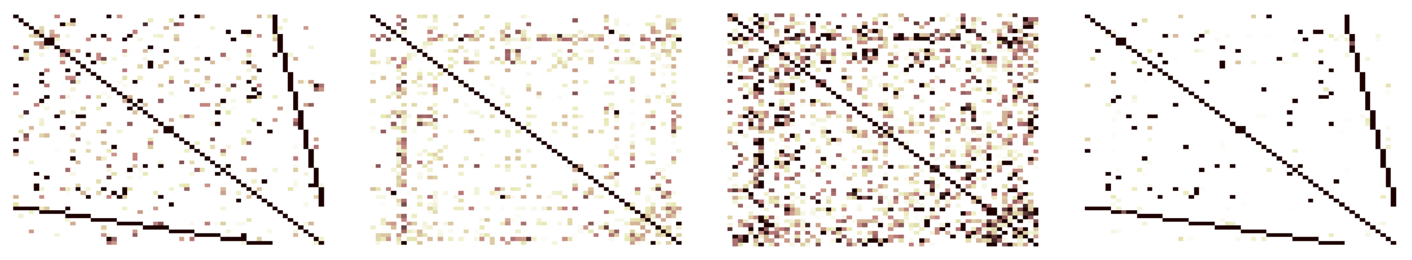
- $\tilde{\Theta}$: **must be similar to the empirical Θ and satisfy sparsity property**
- Similarity measured by KL divergence / Sparsity controlled by $|\tilde{\Theta}|_1$.
- Using the two properties, we formulate our problem as

$$\max_{s\geq 0} \ \mathrm{tr}(\tilde{\Theta}\Theta^{-1}) - \log\det(\tilde{\Theta}\Theta^{-1}) - n + \gamma|\tilde{\Theta}|_1$$

- Substituting $\tilde{\Theta} = \sum_{\ell=1}^{n} \sigma_\ell g^2(s\sigma_\ell)V_\ell V_\ell^T$ yields

$$\max_{s\geq 0} \sum_{\ell=1}^{n} \lambda_\ell K(s,\sigma_\ell) - \sum_{\ell=1}^{n} \log(\lambda_\ell K(s,\sigma_\ell)) - n + \gamma\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sum_{\ell=1}^{n} K(s,\sigma_\ell)X_\ell(i,j)\right| \tag{8}$$

- We solve eq. (8) using gradient descent to obtain the optimal scale s for $\tilde{\Theta}$ that yields **sparsity** and **similarity to** Θ at the same time.

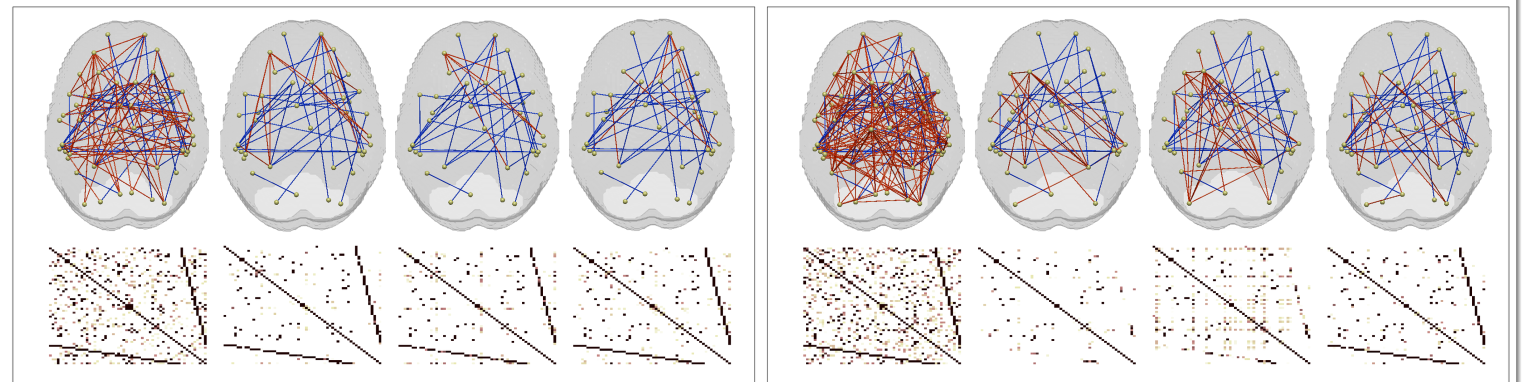


(a) (b) (c) (d)

Figure : Precision matrix in different scales. a) estimation with $s = 0$, b) estimation with $s = 0.5$, c) estimation with $s = 0.7$, d) estimation with optimal scale $s = 0.2089$.

Latent Variable Graphical Model Selection

- Graphical model estimation on a population of synthetic brain pathways.
- Pathways $n_p = 50$, covariates $n_c = 10$ (total observed $n_o = n_p + n_c = 60$).
- Generate data based on a true precision matrix with 5% density.
- Kernel function $g(\sigma) = \sigma e^{-s\sigma}$
- Different number of latent variables $n_h = 5, 10$ to see the effect of latent variables in different methods.



$n_h = 5$ $n_h = 10$

Figure : Comparison of statistical dependency estimations between observed variables (when there are at least a few latent components) using synthetic brain network data.

- Top/bottom rows show estimated dependencies in the data (correct estimation in blue and false positive in red) and their corresponding precision matrices.
- First col: sample precision matrix, Second col: result using GLasso, Third col: result using Chandrasekaran et al, Fourth col: our result.

Experiments on Human Connectome Project Data

- **Imaging Data.** Tractography (structural brain connectivity) derived from Diffusion Tensor Image (DTI) from HCP over 489 participants and 17 major white-matter connectivities selected
- **Non-imaging Data.** 22 non-imaging covariates (out of ~350) related to demographics, physical health, memory, sleep and etc.

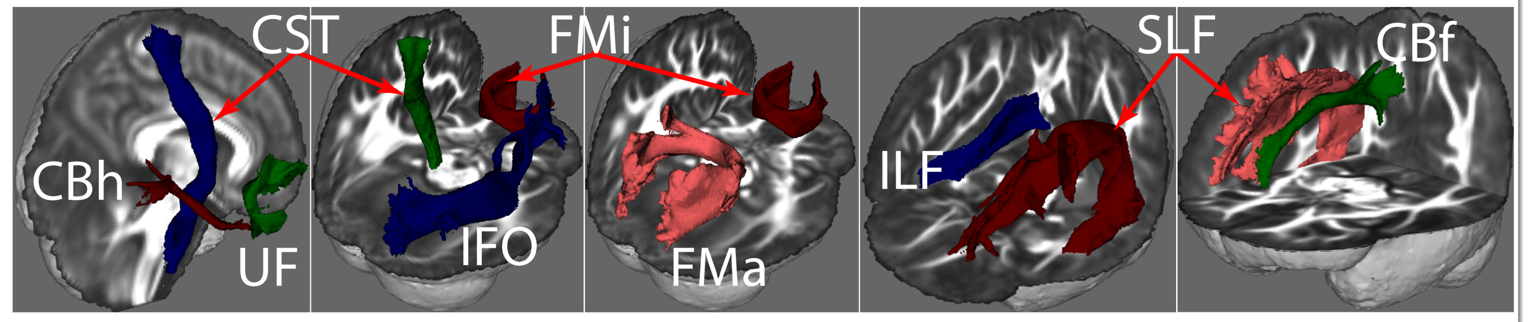


Figure : The 17 connections that are statistically associated to non-imaging covariates.

- **Goal.** Recover a sparse (and interpretable) precision matrix explaining the conditional dependencies among the variables despite latent variables

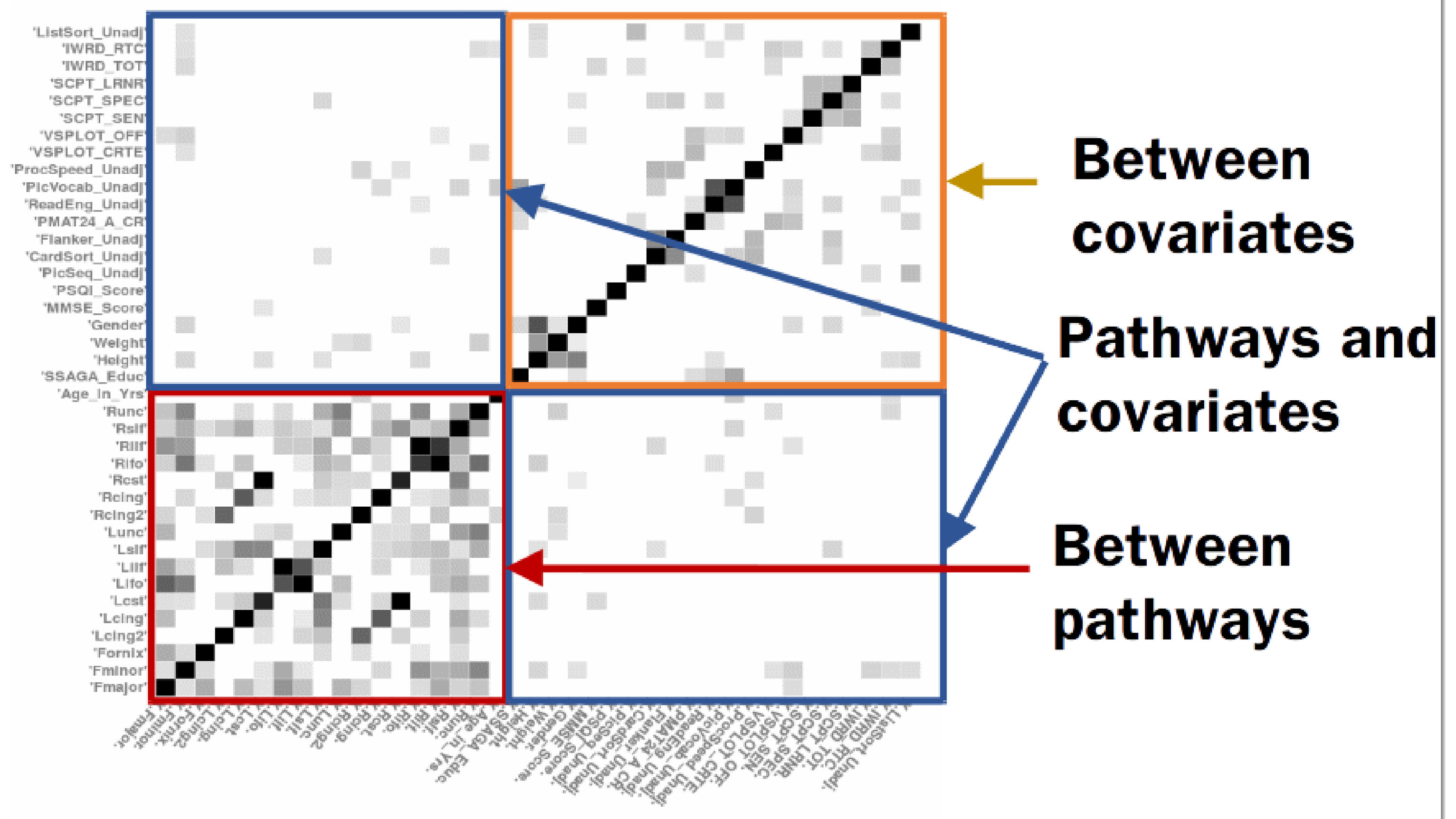


Figure : Estimated precision matrix on HCP dataset.

- **Analysis.** A parsimonious set of relations identified among the non-imaging covariates (in blue) / among the brain pathways (in red) / across these two groups of variables (in orange)
- Major associations identified between **1) cingulum bundle and processing speed, 2) longitudinal fasciculus and cognitive/verbal ability, 3) forceps major and gender, 4) uncinate fasciculus and spatial working memory and etc.**
- The identified relationships are corroborated by results from independent literature.

Acknowledgment

This research was supported by NIH grants AG040396, NSF CAREER award 1252725, UW ADRC AG033514, UW ICTR 1UL1RR025011, UW CPCP AI117924 and Waisman Core Grant P30 HD003352-45. Visit http://cs.wisc.edu/$\sim$wonhwa/project or http://cs.wisc.edu/hwkim for code and supplements.

*: first authorship shared.