Assumption-Lean Inference for Spectral Differential Network Analysis of High-Dimensional Time Series

Michael Hellstern

Department of Biostatistics, University of Washington

Joint work with...

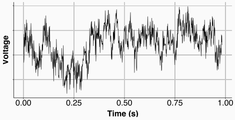
Ali Shojaie
University of Washington

Byol Kim
Sookmyung Women's University

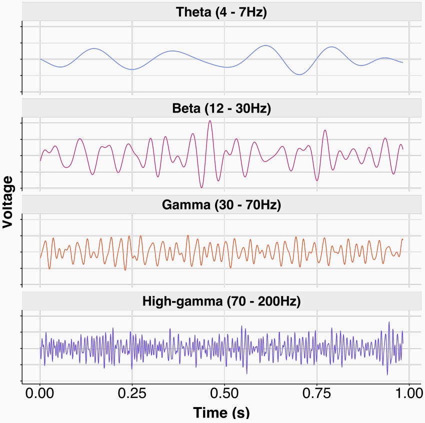
Related work: Hellstern M, Kim B, Harchaoui Z, Shojaie A. *Spectral Differential Network Analysis for High-dimensional Time Series*. AISTATS (2025)

Spectral analysis of time series data

- **Spectral analysis** studies signal at frequency instead of time
- In neuroscience, higher frequency oscillations contain essential brain connectivity information

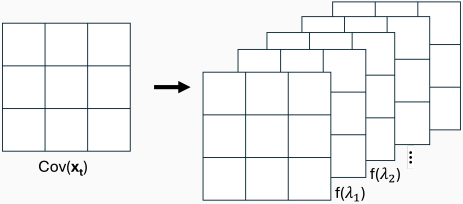


→



Spectral density

- The **spectral density, $f(\lambda)$,** is frequency domain version of covariance
- **Spectral density** decomposes covariance by frequency



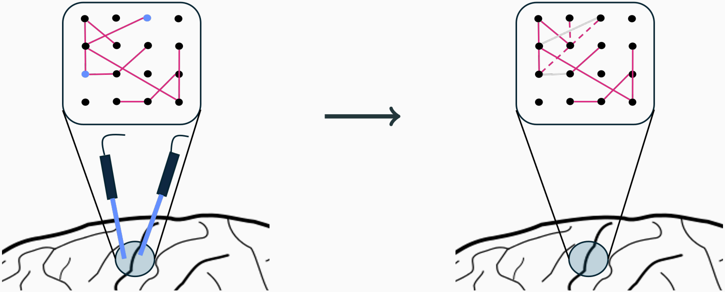
Spectral networks

- **Inverse spectral density**, $f^{-1}(\lambda)$, is a useful network since it encodes conditional relationships[1]
- $f^{-1}(\lambda)_{i,j}$ encodes the (rescaled) coherence between node i and node j after removing the linear effects of all other nodes

[1] Dahlhaus (2000)

Why study a difference in networks?

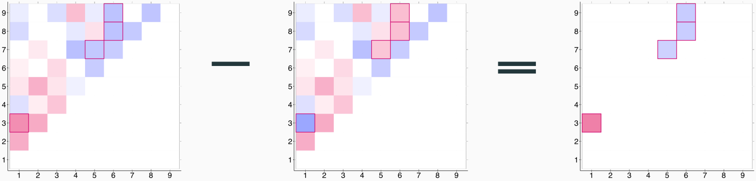
- Studying how external stimuli change the brain network at different frequencies we can hopefully develop treatments



- We are therefore interested in estimating $\mathbf{f}_1^{-1}(\boldsymbol{\lambda}) - \mathbf{f}_2^{-1}(\boldsymbol{\lambda})$

Current methods to estimate a difference in $f^{-1}(\lambda)$

- Only existing method is to estimate $f^{-1}(\lambda)$ in each condition and take difference
- In high-dimensions consistency of $\hat{f}_1^{-1} - \hat{f}_2^{-1}$ requires both $\hat{f}_1^{-1}, \hat{f}_2^{-1}$ to be sparse (unrealistic in practice)



From the complex to real space

- Brief detour to frame problem in real space

$$\Sigma_i = \begin{bmatrix} \mathrm{Re}(f_i) & -\mathrm{Im}(f_i) \\ \mathrm{Im}(f_i) & \mathrm{Re}(f_i) \end{bmatrix}, \quad \Sigma_i^{-1} = \begin{bmatrix} \mathrm{Re}(f_i^{-1}) & -\mathrm{Im}(f_i^{-1}) \\ \mathrm{Im}(f_i^{-1}) & \mathrm{Re}(f_i^{-1}) \end{bmatrix}$$

- $f_1^{-1} - f_2^{-1}$ can easily be recovered from $\Sigma_1^{-1} - \Sigma_2^{-1}$

Spectral D-trace Difference (SDD) Estimator

- Note
$$\frac{\partial L_D}{\partial \Delta}(\Delta) = \left(\Sigma_2 \Delta \Sigma_1 + \Sigma_1 \Delta \Sigma_2\right)/2 - \left(\Sigma_2 - \Sigma_1\right), \tag{1}$$
is estimating equation for $\Delta^* = \Sigma_1^{-1} - \Sigma_2^{-1}$
- We use the integral of $\frac{\partial L_D}{\partial \Delta}(\Delta)^2$ and include a penalty on $\|\Delta\|_1$ to induce sparsity

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \frac{1}{4}\left(\langle \hat{\Sigma}_2 \Delta, \Delta \hat{\Sigma}_1 \rangle + \langle \hat{\Sigma}_1 \Delta, \Delta, \hat{\Sigma}_2 \rangle\right) - \langle \Delta, \hat{\Sigma}_2 - \hat{\Sigma}_1 \rangle + \tau \|\Delta\|_1$$

- $\hat{\Sigma}_i$ are sample estimates of Σ_i - smoothed periodograms with a smoothing bandwidth of B

[2] This is referred to as the D-trace loss and was first introduced in Yuan et al. (2017)

Inference using de-biased estimating equations

- $\hat{\Delta}$ is consistent for Δ under mild data dependence conditions
- Inference can be performed using de-biasing methods for high-dimensional estimating equations[3]
    - De-biases by projecting estimating equation on sparse direction
- Require estimate of $\frac{\partial^2 L_D}{\partial \Delta^2} = \left(\Sigma_1 \otimes \Sigma_2 + \Sigma_2 \otimes \Sigma_1\right)^{-1}$

[3] Neykov et al. (2018)

Inference approach

- Instead of $\text{symm} = (\Sigma_2 \Delta \Sigma_1 + \Sigma_1 \Delta \Sigma_2)/2 - (\Sigma_2 - \Sigma_1)$ we can use either

$$\text{s1Right} = \Sigma_2 \Delta \Sigma_1 - (\Sigma_2 - \Sigma_1) \qquad \text{or}$$
$$\text{s1Left} = \Sigma_1 \Delta \Sigma_2 - (\Sigma_2 - \Sigma_1)$$

- Then inference requires estimates of

$$(\Sigma_1 \otimes \Sigma_2)^{-1} = \Sigma_1^{-1} \otimes \Sigma_2^{-1} \qquad \text{or}$$
$$(\Sigma_2 \otimes \Sigma_1)^{-1} = \Sigma_2^{-1} \otimes \Sigma_1^{-1}$$

Theoretical results

Theorem 1

If the dependence in the data is not too strong[4], for appropriate choices of penalty parameters we have

$$\frac{1}{\hat{\sigma}_{\text{symm}}}\sqrt{\frac{T}{B}}\left(\tilde{\theta}_{\text{symm}}-\theta^*\right)\xrightarrow[T\to\infty]{d}N(0,1)$$
$$\frac{1}{\hat{\sigma}_{\text{s1Left}}}\sqrt{\frac{T}{B}}\left(\tilde{\theta}_{\text{s1Left}}-\theta^*\right)\xrightarrow[T\to\infty]{d}N(0,1)$$
$$\frac{1}{\hat{\sigma}_{\text{s1Right}}}\sqrt{\frac{T}{B}}\left(\tilde{\theta}_{\text{s1Right}}-\theta^*\right)\xrightarrow[T\to\infty]{d}N(0,1)\,,$$

where $\tilde{\theta}$ is the de-biased estimate. B is the bandwidth used to generate sample estimates of spectral density.

[4]details in Appendix

Inference simulations

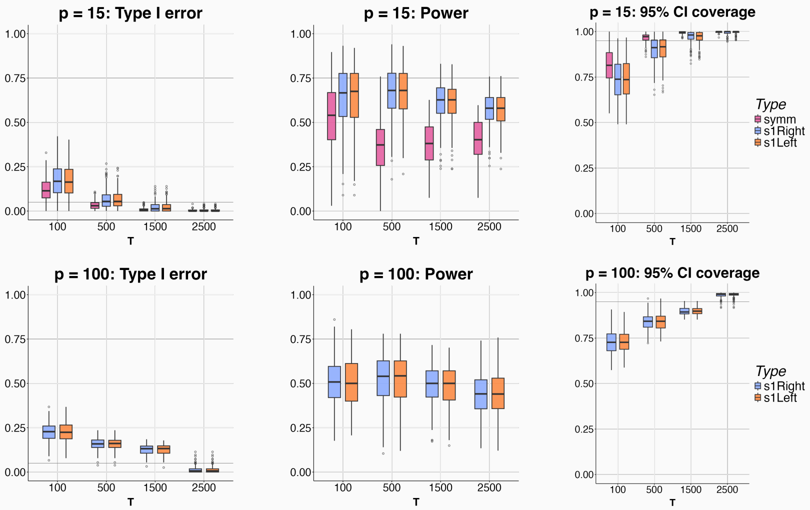


Figure 1: Top row: $p = 15$. Bottom row: $p = 100$

Application to EEG data

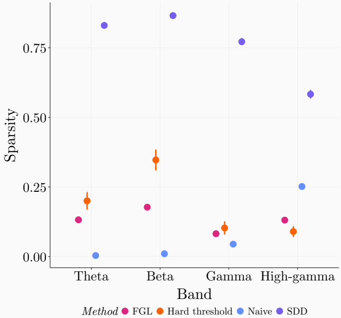
- 64 channel EEG recording of 111 healthy subjects [5]
- 4 minutes recording while subjects at rest with eyes closed
- 42 subjects have follow-up session 2-3 months later

- Subjects with follow-up (across session analysis)
  - Estimated differential network across sessions
- Subjects without follow-up (within session analysis)
  - Estimated differential network from 0-60s to 120-180s

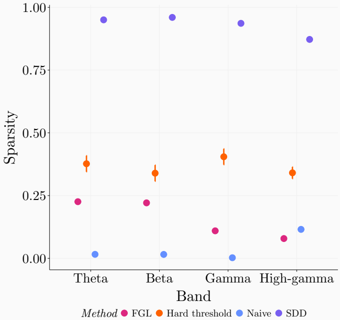
[5] This data is from Hatlestad-Hall et al. (2022)

Application to EEG data

Within session analysis has sparser differential network



Across session analysis



Within session analysis

Questions?

References

Rainer Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51:157–172, 2000.

Christoffer Hatlestad-Hall, Trine Waage Rygvold, and Stein Andersson. Bids-structured resting-state electroencephalography (eeg) data extracted from an experimental paradigm. *Data in Brief*, 45:108647, 2022.

Matey Neykov, Yang Ning, Jun S. Liu, and Han Liu. A Unified Theory of Confidence Regions and Testing for High-Dimensional Estimating Equations. *Statistical Science*, 33(3):427 – 443, 2018. doi: 10.1214/18-STS661. URL `https://doi.org/10.1214/18-STS661`.

Huili Yuan, Ruibin Xi, Chong Chen, and Minghua Deng. Differential network analysis via lasso penalized d-trace loss. *Biometrika*, 104(4): 755–770, 2017.

Chi Zhang and Danna Zhang. Spectral inference for high dimensional time series. *IEEE Transactions on Information Theory*, 2025.

Appendix

Spectral density definition

Let $\mathbf{x}_t \in \mathbb{R}^{p\times 1}$ then autocovariance at lag $h \in \mathbb{Z}$ is denoted as $\Gamma(h) = \mathbb{E}\left(\mathbf{x}_t \mathbf{x}_{+h}^T\right)$. **Spectral density, $\mathbf{f}(\lambda)$,** at frequency λ is then defined as

$$f(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-i\lambda h}\Gamma(h), \qquad -\pi \leq \lambda \leq \pi.$$

Periodogram and smoothing

For condition l, the periodogram at Fourier frequency $\{\lambda_j = 2\pi j/n_l, -\lfloor (n_l - 1)/2 \rfloor \leq j \leq \lfloor n_l/2 \rfloor\}$ is defined as

$$P_l(\lambda_j) = \frac{1}{2\pi n_l} \left(\sum_{t=1}^{n_l} \mathbf{x}_{l,t} \exp(-i\lambda_j t) \right) \left(\sum_{t=1}^{n_l} \mathbf{x}_{l,t} \exp(-i\lambda_j t) \right)^H .$$

But periodogram not consistent, so must smooth over nearby frequencies

$$\hat{f}_l(\lambda_j) = \frac{1}{2B+1} \sum_{k=j-B}^{j+B} P_l(\lambda_k).$$

Functional dependence

Let ϵ_0^* and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be i.i.d. vectors in $\mathbb{R}^b$ and $\mathcal{F}_t = (\ldots, \epsilon_{t-1}, \epsilon_t)$. We further let $\mathbf{x}_{l,t} = (x_{l,1t}, \ldots x_{l,pt})^T$ be a p-dimensional process in condition l where

$$x_{l,jt} = R_{l,j}(\mathcal{F}_t).$$

To measure the dependence on ϵ_0 we replace ϵ_0 in $\mathcal{F}_t$ with an i.i.d. copy ϵ_0^*. This gives $\mathcal{F}_t' = (\ldots, \epsilon_{-1}, \epsilon_0^*, \epsilon_1, \ldots, \epsilon_t)$ and

$$x_{l,jt}' = R_{l,j}(\mathcal{F}_t').$$

The dependence on ϵ_0 can then be measured by

$$\delta_{l,t,q,j} = \left(\mathbb{E}\left|x_{l,jt} - x_{l,jt}'\right|^q\right)^{1/q}$$

where $\mathbb{E}\,|x|^q$ is the expected value of $|x|^q$.

Assume there exists $\rho \in (0,1)$ such that

$$\|X_{l,\cdot j}\|_{L_q} = \sup_{m\geq 0} \rho^{-m} \Delta_{l,m,q,j} < \infty, \text{where } \Delta_{l,m,q,j} = \sum_{t=m}^{\infty} \delta_{l,t,q,j}\,.$$

Dependence assumption

Assumption 1

There exists some constant $\alpha_l \geq 0$ *such that*

$$\|X_{l,\cdot j}\|_{\psi_{\alpha_l}} := \sup_{q \geq 2} q^{-\alpha_l} \|X_{l,\cdot j}\|_{L_q} < \infty \, .$$

In Assumption 1, α_l represents the dependence in the process for condition l. As noted in Zhang and Zhang (2025), in the iid case, $\alpha_l = 1/2$ corresponds to the classical sub-Gaussian norm while $\alpha_l = 1$ corresponds to the sub-Exponential norm. The larger α_l is, the heavier the tails.

Form of asymptotic variance of de-biased SDD

The full asymptotic distribution of the (symm) de-biased D-trace loss function is

$$\sqrt{T/B}v^{*T}\mathrm{vec}\left[\left(\hat{\Sigma}_2\Delta^*\hat{\Sigma}_1+\hat{\Sigma}_1\Delta^*\hat{\Sigma}_2\right)/2-\left(\hat{\Sigma}_2-\hat{\Sigma}_1\right)\right]\rightsquigarrow N\left(0,v^{*T}\left(M_1V_1M_1^T+M_2V_2M_2^T\right)v^*\right)$$

where

- Δ^* is the true difference
- $v^* = \left(\frac{\partial^2 L_D}{\partial \Delta^2}^{-1}\right)_{i*}$
- $M_1 = \frac{\left(\Sigma_2^T\Delta^{*,T}\otimes I\right)+(I\otimes\Sigma_2\Delta^*)}{2}+I$
- $M_2 = \frac{\left(\Sigma_1^T\Delta^{*,T}\otimes I\right)+(I\otimes\Sigma_1\Delta^*)}{2}-I$
- $V_i = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} V_i' \begin{bmatrix} P_1^T & P_2^T \end{bmatrix}$, P_i are known 0/1 permutation matrices
- $$V_i' = \frac{1}{2}\begin{bmatrix} \left(I_{p^2}+K\right)\left(A_i\otimes A_i+B_i\otimes B_i\right) & \left(I_{p^2}+K\right)\left(B_i\otimes A_i-A_i\otimes B_i\right) \\ \left[\left(I_{p^2}+K\right)\left(B_i\otimes A_i-A_i\otimes B_i\right)\right]^T & \left(I_{p^2}-K\right)\left(A_i\otimes A_i+B_i\otimes B_i\right) \end{bmatrix}$$

Inference challenges

1. Requires new concentration inequality on smoothed spectral density estimates
2. Extended de-biasing methods to arbitrary scaling s_n
3. Solve for form of asymptotic variance of $\mathrm{vec}\left(\hat{\Sigma}_i - \Sigma_i\right)$
4. Inference requires estimate of $\frac{\partial^2 L_D}{\partial \Delta^2} = \left(\Sigma_1 \otimes \Sigma_2 + \Sigma_2 \otimes \Sigma_1\right)^{-1}$, too large to compute even for moderate p
5. Moreover, computing asymptotic variance requires multiplication of matrices too large to compute