Regularized Estimation of Sparse Spectral Precision Matrices

Spring Research Conference 2025

Navonil Deb
(Joint work with Amy Kuceyeski and Sumanta Basu)

June 3, 2025

Department of Statistics and Data Science
Cornell University

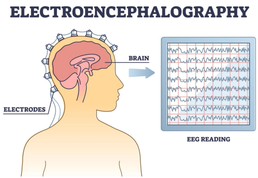
Outline

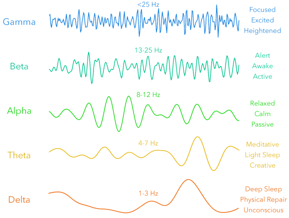
- Frequency-specific graphical model.
- Penalized whittle likelihood estimator of spectral precision matrix.
- Develop a fast, pathwise coordinate descent based estimation procedure.
    1. Complex lasso (CLASSO).
    2. Complex graphical lasso (CGLASSO).
- CGLASSO with adaptive penalization.

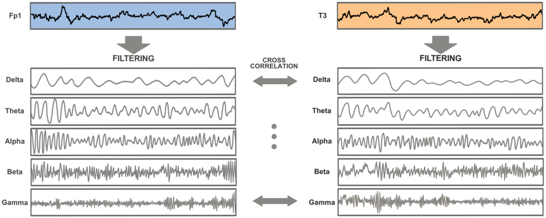
Frequency Domain Graphical Models

Motivation: Construct functional connectivity (FC) network among brain regions.

- Can we learn the *correlation structure* among frequency-*specific* physiological activities?







Reference: Ombao and Pinto (2022).

Frequency-Domain Graphical Models

Stationary, zero-mean $X_t \in \mathbb{R}^p$, autocovariance $\Gamma(h) = \mathbb{E}[X_{t+h}X_t^\top]$.
Spectral density: $f(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-ih\omega}$.

Structure	**Classical Domain**	**Spectral Domain**
Second-order structure	Σ	$f(\omega)$
Conditional dependence	Σ^{-1}	$\Theta(\omega) = f(\omega)^{-1}$
Marginal association	Correlation	Coherence
Conditional association	Partial correlation	Partial coherence

Frequency-Domain Graphical Models

Stationary, zero-mean $X_t \in \mathbb{R}^p$, autocovariance $\Gamma(h) = \mathbb{E}[X_{t+h} X_t^\top]$.
Spectral density: $f(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-ih\omega}$.

Structure	**Classical Domain**	**Spectral Domain**
Second-order structure	Σ	$f(\omega)$
Conditional dependence	Σ^{-1}	$\Theta(\omega) = f(\omega)^{-1}$
Marginal association	Correlation	Coherence
Conditional association	Partial correlation	Partial coherence

- **Applications:** Neuroscience, economics, finance.
  - **Motivation:** Rich description of dependence.
  - Captures contemporaneous and lead-lag associations.
- **Graphical interpretation:** Edge $r \leftrightarrow s$ exists if $[\Theta(\omega)]_{r,s} \neq 0$ for some ω.

Frequency-Domain Graphical Models

Stationary, zero-mean $X_t \in \mathbb{R}^p$, autocovariance $\Gamma(h) = \mathbb{E}[X_{t+h} X_t^\top]$.
Spectral density: $f(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-ih\omega}$.

Structure	**Classical Domain**	**Spectral Domain**
Second-order structure	Σ	$f(\omega)$
Conditional dependence	Σ^{-1}	$\Theta(\omega) = f(\omega)^{-1}$
Marginal association	Correlation	Coherence
Conditional association	Partial correlation	Partial coherence

- **Applications:** Neuroscience, economics, finance.
  - **Motivation:** Rich description of dependence.
  - Captures contemporaneous and lead-lag associations.
- **Graphical interpretation:** Edge $r \leftrightarrow s$ exists if $[\Theta(\omega)]_{r,s} \neq 0$ for some ω.
- **Goal:** Estimate full $\Theta(\omega)$ at given ω, rather than single entry across all ω.
  – **Challenges**: Heteroskedasticity, high-dimensionality, temporal dependence.

References: Dahlhaus (2000), Böhm and von Sachs (2009)

Bridging the Gap: Our Contributions

Limitations of existing approaches.

1. Off-the-shelf solvers e.g. ADMM do not exploit the problem structure, not adaptive. (Baek et al., 2021)
2. Real imaginary separation $\implies$ dimension $\times 4$. (Fiecas et al., 2019, SIPE based on CLIME).
3. Separate treatment of real and imaginary parts in nodewise regression ignores joint likelihood. (Krampe and Paparoditis, 2022)

Bridging the Gap: Our Contributions

Limitations of existing approaches.

1. Off-the-shelf solvers e.g. ADMM do not exploit the problem structure, not adaptive. (Baek et al., 2021)
2. Real imaginary separation $\implies$ dimension $\times 4$. (Fiecas et al., 2019, SIPE based on CLIME).
3. Separate treatment of real and imaginary parts in nodewise regression ignores joint likelihood.
   (Krampe and Paparoditis, 2022)

Our contributions.

- **Algorithmic**: CGLASSO — fast, pathwise coordinate descent-based algorithm.
  – Leverages high-dimensional tricks viz. warm starts and active set screening.
- **Optimization Insight**: Ring isomorphism-based realification.
  – Generalizes to broader complex-valued problems.

– **Methodological**: Adaptive CGLASSO introduces entry-wise penalization.
  – Improved structural recovery in heterogeneous variability.

– **Theoretical**: Non-asymptotic error bounds for both standard and adaptive CGLASSO.

Averaged Periodogram

- Data: n observations from a p-dim stationary series $\{X_1, X_2, \ldots, X_n\}$.
- Fourier frequencies: $\omega_j = 2\pi j/n$, $j \in F_n = \{-[(n-1)/2], \ldots, [n/2]\}$.
- DFT: $d_j = n^{-1/2} \sum_{t=1}^{n} X_t e^{-it\omega_j}$.

$$X = \begin{bmatrix} X_{11} & \cdots & X_{np} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix} \xrightarrow{\text{FFT}} \begin{bmatrix} d_{j-m,1} & \cdots & d_{j-m,p} \\ \vdots & \ddots & \vdots \\ \color{red}{d_{j,1}} & \color{red}{\cdots} & \color{red}{d_{j,p}} \\ \vdots & \ddots & \vdots \\ d_{j+m,1} & \cdots & d_{j+m,p} \end{bmatrix} = \mathcal{Z}$$

Averaged Periodogram

- Data: n observations from a p-dim stationary series $\{X_1, X_2, \ldots, X_n\}$.
- Fourier frequencies: $\omega_j = 2\pi j/n$, $j \in F_n = \{-[(n-1)/2], \ldots, [n/2]\}$.
- DFT: $d_j = n^{-1/2} \sum_{t=1}^{n} X_t e^{-it\omega_j}$.

$$X = \begin{bmatrix} X_{11} & \cdots & X_{np} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix} \xrightarrow{\text{FFT}} \begin{bmatrix} d_{j-m,1} & \cdots & d_{j-m,p} \\ \vdots & \ddots & \vdots \\ d_{j,1} & \cdots & d_{j,p} \\ \vdots & \ddots & \vdots \\ d_{j+m,1} & \cdots & d_{j+m,p} \end{bmatrix} = \mathcal{Z}$$

- Under regularity conditions, $\mathrm{Cov}(d_{j,r}, d_{j,s}) \approx 2\pi[f(\omega_j)]_{r,s}$ and d_j are asymptotically independent across j.
- Averaged periodogram for bandwidth m: $\hat{f}(\omega_j) = \frac{1}{2\pi(2m+1)} \mathcal{Z}^\top \overline{\mathcal{Z}}$.

Penalized Whittle log-Likelihood

- For Gaussian time series, $d_j \sim \mathcal{N}_{\mathbb{C}}(0, 2\pi\Theta^{-1}(\omega_j))$, asymp. independent across j.
- Approximated Whittle (negative log) likelihood takes the form

$$\frac{1}{2} \sum_{j \in F_n} \log\det \Theta(\omega_j) - \sum_{j \in F_n} d_j^* \, \Theta(\omega_j) \, d_j$$

Penalized Whittle log-Likelihood

- For Gaussian time series, $d_j \sim \mathcal{N}_{\mathbb{C}}(0, 2\pi\Theta^{-1}(\omega_j))$, asymp. independent across j.
- Approximated Whittle (negative log) likelihood takes the form

$$\frac{1}{2}\sum_{j\in F_n} \log\det\Theta(\omega_j) - \sum_{j\in F_n} d_j^* \, \Theta(\omega_j)\, d_j$$

- Approximated and penalized problem:

$$\boxed{\hat{\Theta}_j \equiv \hat{\Theta}(\omega_j) = \underset{\Theta\in\mathcal{H}^p_{++}}{\text{argmin}} \; -\log\det\Theta + \text{trace}(\hat{f}(\omega_j)\Theta) + \lambda\|\Theta\|_{1,\text{off}},}$$

where $\|\Theta\|_{1,\text{off}} = \sum_{r\neq s}|\Theta_{rs}|$.

– We call this problem **CGLASSO** $\equiv$ complex graphical lasso.

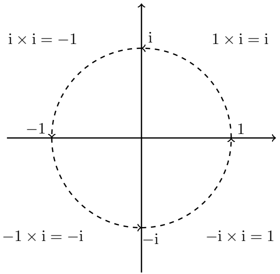
The Isomorphism Trick

We consider the map $\varphi(z) : \mathbb{C} \to \mathbb{R}^{2\times 2}$ s.t.

$$\varphi(z) = \begin{bmatrix} \mathbf{Re}(z) & -\mathbf{Im}(z) \\ \mathbf{Im}(z) & \mathbf{Re}(z) \end{bmatrix}.$$

- Properties of $\varphi(z) = (\varphi_1(z), \varphi_2(z))$:

1. $\langle \varphi_1(z), \varphi_2(z) \rangle = 0$.
2. $\|\varphi_1(z)\|^2 = \|\varphi_2(z)\|^2 = |z|^2$.
3. $\varphi(z^\dagger) = [\varphi(z)]^\top$.
4. φ is a field isomorphism.



$$1 \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathrm{i} \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

1 and $\mathrm{i} \equiv 2 \times 2$ orthogonal matrices!

The Isomorphism Trick

φ: extendable to $\mathbb{C}^p$ and $\mathbb{C}^{n\times p}$ with a k-dim permutation matrix Π_k.

'Tilde' notation:

$$\begin{aligned} \tilde{z} &= \Pi_n^\top \varphi(z) e_1 = \begin{pmatrix} \mathbf{Re}(z) \\ \mathbf{Im}(z) \end{pmatrix}, \\ \tilde{\tilde{Z}} &= \Pi_m^\top \varphi(Z) \Pi_n = \begin{pmatrix} \mathbf{Re}(z) & -\mathbf{Im}(Z) \\ \mathbf{Im}(Z) & \mathbf{Re}(Z) \end{pmatrix}. \end{aligned}$$

The Isomorphism Trick

φ: extendable to $\mathbb{C}^p$ and $\mathbb{C}^{n\times p}$ with a k-dim permutation matrix Π_k.

'Tilde' notation:

$$\tilde{z} = \Pi_n^\top \varphi(z) e_1 = \begin{pmatrix} \mathbf{Re}(z) \\ \mathbf{Im}(z) \end{pmatrix},$$

$$\tilde{\tilde{Z}} = \Pi_m^\top \varphi(Z) \Pi_n = \begin{pmatrix} \mathbf{Re}(z) & -\mathbf{Im}(Z) \\ \mathbf{Im}(Z) & \mathbf{Re}(Z) \end{pmatrix}.$$

Context	Complex Loss Function	Realified Loss Function
OLS (X, y, β)	$\frac{1}{2n}\lVert y - X\beta\rVert^2$	$\frac{1}{4n}\lVert \tilde{y} - \tilde{\tilde{X}}\tilde{\beta}\rVert^2$
log-det program $Z_i \sim \mathcal{N}(0, \Theta^{-1})$ $P = \frac{1}{n}\sum_i Z_i Z_i^\dagger$	$-\log\det\Theta + \mathrm{trace}(P\Theta)$	$\frac{1}{2}\left[-\log\det\tilde{\tilde{\Theta}} + \mathrm{trace}(\tilde{\tilde{P}}\tilde{\tilde{\Theta}})\right]$

CLASSO $\equiv$ Group Lasso

Complex Lasso (CLASSO):

$$\underset{\beta \in \mathbb{C}^p}{\operatorname{argmin}} \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1.$$

CLASSO $\equiv$ Group Lasso

Complex Lasso (CLASSO):

$$\underset{\beta \in \mathbb{C}^p}{\operatorname{argmin}} \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1.$$

$$\Updownarrow$$

$$\boxed{\underset{\tilde{\beta}_1, \ldots, \tilde{\beta}_p \in \mathbb{R}^2}{\operatorname{argmin}} \frac{1}{2n} \Big\| \tilde{Y} - \sum_{j=1}^{p} \tilde{\tilde{X}}_j \tilde{\beta}_j \Big\|_2^2 + \lambda \sum_{j=1}^{p} \|\tilde{\beta}_j\|_2.}$$

CLASSO $\equiv$ Group Lasso

Complex Lasso (CLASSO):

$$\underset{\beta\in\mathbb{C}^p}{\operatorname{argmin}} \frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1.$$

$$\Updownarrow$$

$$\boxed{\underset{\tilde\beta_1,\ldots,\tilde\beta_p\in\mathbb{R}^2}{\operatorname{argmin}} \frac{1}{2n}\Big\|\tilde{Y} - \sum_{j=1}^{p}\tilde{\tilde{X}}_j\tilde{\beta}_j\Big\|_2^2 + \lambda\sum_{j=1}^{p}\|\tilde{\beta}_j\|_2.}$$

- CLASSO $\equiv$ Group Lasso over reals Maleki et al. (2013)
  - Orthogonal predictors within a group.
  - Closed form updates of (blockwise) coordinate descent.

CGLASSO Algorithm Sketch

$\hat{f} \equiv \hat{f}(\omega_j)$ for a fixed ω_j.

$$\hat{\Theta}_j = \underset{\Theta \in \mathcal{H}^p_{++}}{\operatorname{argmin}} \; -\log\det\Theta + \operatorname{trace}(\hat{f}\Theta) + \lambda\|\Theta\|_{1,\text{off}}.$$

Along Friedman et al. (2008):

- **KKT condition**: $\boxed{\hat{f} - \Theta^{-1} + \lambda\Psi = 0}$, $\Psi_{r,s} = \operatorname{sign}(\Theta_{r,s})$.

CGLASSO Algorithm Sketch

$\hat{f} \equiv \hat{f}(\omega_j)$ for a fixed ω_j.

$$\hat{\Theta}_j = \underset{\Theta \in \mathcal{H}^p_{++}}{\operatorname{argmin}} \; -\log\det\Theta + \operatorname{trace}(\hat{f}\Theta) + \lambda\|\Theta\|_{1,\text{off}}.$$

Along Friedman et al. (2008):

- **KKT condition**: $\boxed{\hat{f} - \Theta^{-1} + \lambda\Psi = 0}$, $\Psi_{r,s} = \operatorname{sign}(\Theta_{r,s})$.
- W: Working version of Θ^{-1}.
- Partition Θ as $\begin{bmatrix} \Theta_{11} & \boldsymbol{\theta}_{12} \\ \boldsymbol{\theta}^{\dagger}_{12} & \theta_{22} \end{bmatrix}$, same for W and $\hat{f}$.
- KKT + $[\Theta W = I_p] \implies$ Last column iteration :

$$W_{11}\beta - \hat{f}_{12} + \lambda \operatorname{sign}(\beta) = 0, \;\; \beta = -\boldsymbol{\theta}_{12}/\theta_{22}.$$

CGLASSO Algorithm Sketch

$\hat{f} \equiv \hat{f}(\omega_j)$ for a fixed ω_j.

$$\hat{\Theta}_j = \underset{\Theta \in \mathcal{H}^p_{++}}{\operatorname{argmin}} \; -\log\det\Theta + \operatorname{trace}(\hat{f}\Theta) + \lambda\|\Theta\|_{1,\text{off}}.$$

Along Friedman et al. (2008):

- **KKT condition**: $\boxed{\hat{f} - \Theta^{-1} + \lambda\Psi = 0}$, $\Psi_{r,s} = \operatorname{sign}(\Theta_{r,s})$.
- W: Working version of Θ^{-1}.
- Partition Θ as $\begin{bmatrix} \Theta_{11} & \boldsymbol{\theta}_{12} \\ \boldsymbol{\theta}_{12}^{\dagger} & \theta_{22} \end{bmatrix}$, same for W and $\hat{f}$.
- KKT + $[\Theta W = I_p] \implies$ Last column iteration :

$$W_{11}\beta - \hat{f}_{12} + \lambda\,\operatorname{sign}(\beta) = 0, \;\; \beta = -\boldsymbol{\theta}_{12}/\theta_{22}.$$

- KKT condition of CLASSO(X, y, β):

$$\left(\frac{1}{n}X^{\dagger}X\right)\beta - \left(\frac{1}{n}X^{\dagger}y\right) + \lambda\,\operatorname{sign}(\beta) = 0.$$

Pathwise performance enhancement

Regularization grid: $\{\lambda_t\}$.

Warm start (CGLASSO)
$\hat{\Theta}(\lambda_{t-1}) \rightarrow \hat{\Theta}(\lambda_t)$

+

Warmer start (CLASSO)
$\hat{\beta}^{(\text{ cycle}=k-1)} \rightarrow \hat{\beta}^{(\text{ cycle}=k})$

Pathwise performance enhancement

Regularization grid: $\{\lambda_t\}$.

Warm start (CGLASSO)

$\hat{\Theta}(\lambda_{t-1}) \to \hat{\Theta}(\lambda_t)$

+

Warmer start (CLASSO)

$\hat{\beta}^{(\text{ cycle}=k-1)} \to \hat{\beta}^{(\text{ cycle}=k)}$

- Sequential strong *active set screening* (Tibshirani et al., 2012).
- **Convergence safeguard**: *Early stopping rule* implementation based on the underlying selection criterion (e.g. estimation error, BIC etc).
  – Prevents unstable updates at small λ.

Adaptive CGLASSO

- **CGLASSO-1**: Coherency scaling-

$$\hat{D}^2 = \mathrm{diag}(\hat{f}_{1,1}, \ldots, \hat{f}_{p,p}).$$

$$\hat{\Theta}_D \leftarrow \hat{D}^{-1} \cdot \text{CGLASSO}(\hat{D}^{-1}\hat{f}\hat{D}^{-1}, \lambda) \cdot \hat{D}^{-1}.$$

  – Similar to Janková and van de Geer (2018).

Adaptive CGLASSO

- **CGLASSO-1**: Coherency scaling-

$$\hat{D}^2 = \text{diag}(\hat{f}_{1,1}, \ldots, \hat{f}_{p,p}).$$

$$\hat{\Theta}_D \leftarrow \hat{D}^{-1} \cdot \text{CGLASSO}(\hat{D}^{-1}\hat{f}\hat{D}^{-1}, \lambda) \cdot \hat{D}^{-1}.$$

  – Similar to Janková and van de Geer (2018).

- **CGLASSO-2**: CLASSO level normalization-

$$(p-1)\text{-dim diagonal matrix } \hat{D}_{11} \text{ with } (\hat{D}^2_{11})_{r,r} = (\hat{f}_{11})_{r,r}.$$

$$\hat{\beta} \leftarrow \hat{D}_{11}^{-1} \cdot \text{CLASSO}(\hat{D}_{11}^{-1}\hat{f}_{11}\hat{D}_{11}^{-1},\ \hat{D}_{11}^{-1}\hat{f}_{12},\ \lambda) \cdot \hat{D}_{11}^{-1}.$$

Adaptive CGLASSO

- **CGLASSO-1**: Coherency scaling-

$$\hat{D}^2 = \text{diag}(\hat{f}_{1,1}, \ldots, \hat{f}_{p,p}).$$

$$\hat{\Theta}_D \leftarrow \hat{D}^{-1} \cdot \text{CGLASSO}(\hat{D}^{-1} \hat{f} \hat{D}^{-1}, \lambda) \cdot \hat{D}^{-1}.$$

  – Similar to Janková and van de Geer (2018).

- **CGLASSO-2**: CLASSO level normalization-

$$(p-1)\text{-dim diagonal matrix } \hat{D}_{11} \text{ with } (\hat{D}_{11}^2)_{r,r} = (\hat{f}_{11})_{r,r}.$$

$$\hat{\beta} \leftarrow \hat{D}_{11}^{-1} \cdot \text{CLASSO}(\hat{D}_{11}^{-1} \hat{f}_{11} \hat{D}_{11}^{-1},\ \hat{D}_{11}^{-1} \hat{f}_{12},\ \lambda) \cdot \hat{D}_{11}^{-1}.$$

- **CGLASSO-3**: NWR partial variance scaling-

$$\hat{\tau}_k^2 : \text{Estimated MSE from CLASSO}(\mathcal{Z}_{-k}, \mathcal{Z}_k, \lambda_k).$$

$$\hat{D}^2 = \text{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2).$$

$$\hat{\Theta}_D \leftarrow \hat{D}^{-1} \cdot \text{CGLASSO}(\hat{D}^{-1} \hat{f} \hat{D}^{-1}, \lambda) \cdot \hat{D}^{-1}.$$

Adaptive CGLASSO

- **CGLASSO-1**: Coherency scaling-

$$\hat{D}^2 = \mathrm{diag}(\hat{f}_{1,1}, \ldots, \hat{f}_{p,p}).$$

$$\hat{\Theta}_D \leftarrow \hat{D}^{-1} \cdot \mathrm{CGLASSO}(\hat{D}^{-1}\hat{f}\hat{D}^{-1}, \lambda) \cdot \hat{D}^{-1}.$$

  – Similar to Janková and van de Geer (2018).

- **CGLASSO-2**: CLASSO level normalization-

$$(p-1)\text{-dim diagonal matrix } \hat{D}_{11} \text{ with } (\hat{D}^2_{11})_{r,r} = (\hat{f}_{11})_{r,r}.$$

$$\hat{\beta} \leftarrow \hat{D}_{11}^{-1} \cdot \mathrm{CLASSO}(\hat{D}_{11}^{-1}\hat{f}_{11}\hat{D}_{11}^{-1},\ \hat{D}_{11}^{-1}\hat{f}_{12},\ \lambda) \cdot \hat{D}_{11}^{-1}.$$

- **CGLASSO-3**: NWR partial variance scaling-

$$\hat{\tau}_k^2 : \text{Estimated MSE from } \mathrm{CLASSO}(\mathcal{Z}_{-k}, \mathcal{Z}_k, \lambda_k).$$

$$\hat{D}^2 = \mathrm{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2).$$

$$\hat{\Theta}_D \leftarrow \hat{D}^{-1} \cdot \mathrm{CGLASSO}(\hat{D}^{-1}\hat{f}\hat{D}^{-1}, \lambda) \cdot \hat{D}^{-1}.$$

Method	CGLASSO-1	CGLASSO-2	CGLASSO-3
Scales	$\mathrm{diag}(\hat{f})$	$\mathrm{diag}(\hat{f}_{11})$	$\{\hat{\tau}_k^2\}_{k\in[p]}$

Simulation studies

- Results reported for $\omega_j = 0$ and 50 Monte Carlo trials.
- λ selected with BIC.
- $m \asymp \sqrt{n}$ (Böhm and von Sachs, 2009).

Simulation studies

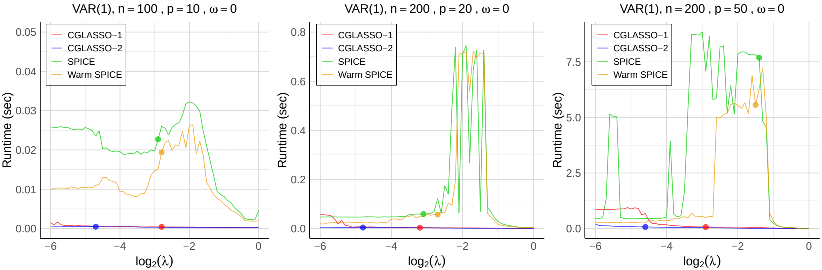
- Results reported for $\omega_j = 0$ and 50 Monte Carlo trials.
- λ selected with BIC.
- $m \asymp \sqrt{n}$ (Böhm and von Sachs, 2009).

Benchmarks:

- **Speed**: *ADMM* vs *coordinate descent*.
- **Estimation accuracy**: SIPE (CLIME) (Fiecas et al., 2019) with *separate penalty* vs CGLASSO with *joint penalty*.
- **Model selection**: *Nodewise regression* of DFTs (Krampe and Paparoditis, 2022) vs *Whittle likelihood* maximization.
- **Scaling**: *Vanilla CGLASSO* vs *adaptive CGLASSO* variants.

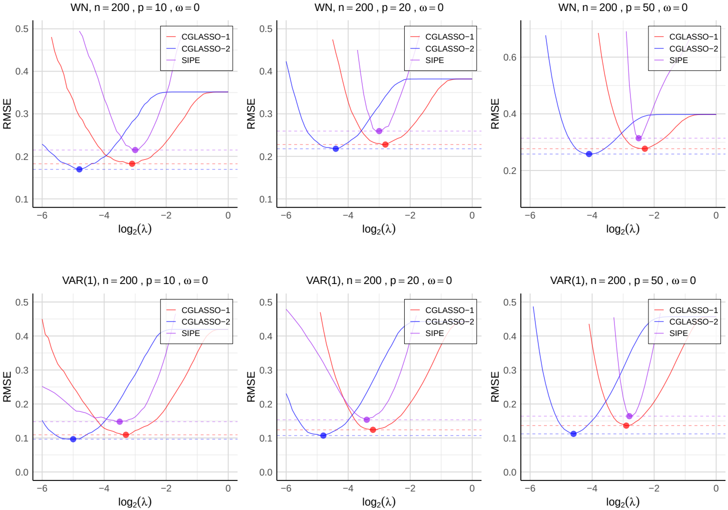
Coordinate descent improves runtime over ADMM

- **SPICE**: Sparse inverse covariance estimator (Scheinberg et al., 2010; Baek et al., 2021), implements ALM algorithm, a variant of ADMM.
- **Warm SPICE**: Pathwise SPICE equipped with warm start.



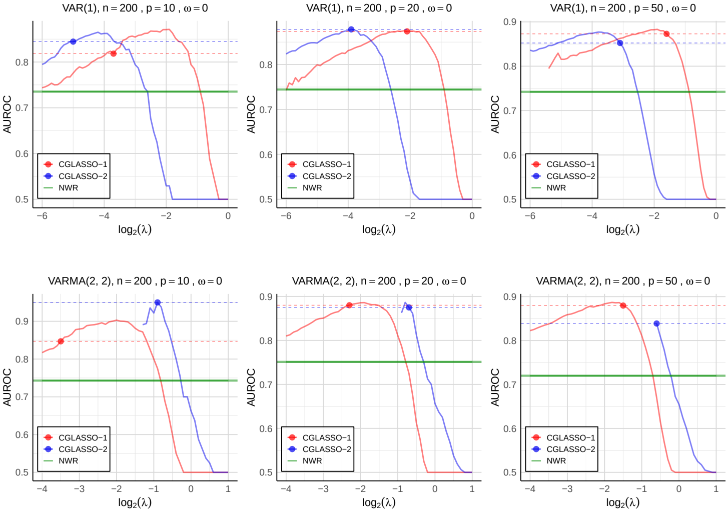
Joint penalty improves estimation accuracy

Benchmark: Sparse inverse periodogram estimator (Fiecas et al., 2019, SIPE), uses CLIME (Cai et al., 2011).



Joint penalty improves model selection

Benchmark: Nodewise regression (Meinshausen and Bühlmann, 2006, NWR) of DFTs (Krampe and Paparoditis, 2022).



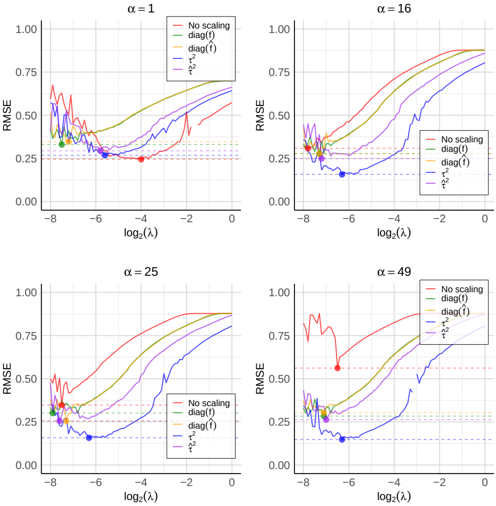
Adaptive penalization improves estimation accuracy

- **DGP**: WN(Σ), Σ^{-1} has two Toeplitz blocks A and B with

$$A_{i,j} = (0.2)^{|i-j|},$$
$$B_{i,j} = (0.8)^{|i-j|}.$$

- α: Signal contrast between the blocks.

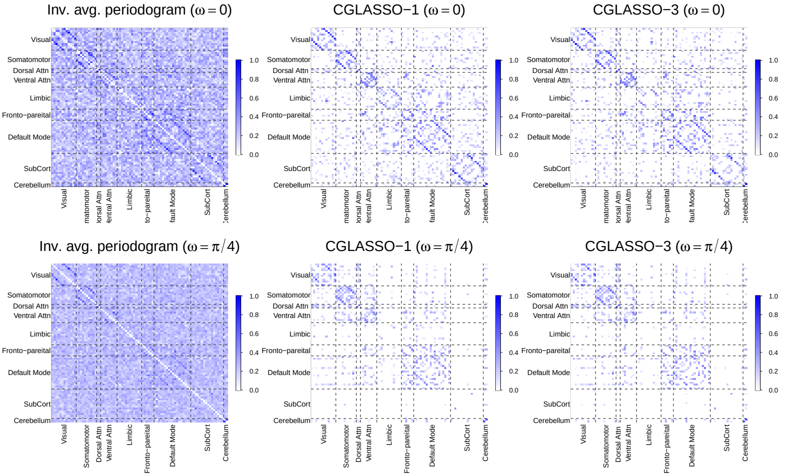
$$\Sigma^{-1} = \frac{1}{15}\begin{bmatrix} A & 0 \\ 0 & \alpha B \end{bmatrix}.$$



fMRI data analysis

CGLASSO applied on resting state fMRI data set with 86 brain region parcellation, from HCP-Young Adult S1200.

Yeo Parcellation Heatmaps (sq-root transformed)



Theoretical guarantees

Bound on estimation error $\|\hat{\Theta}(\omega_j) - \Theta^*(\omega_j)\|_\infty$.

1. *Entry-wise max norm guarantees* for vanilla CGLASSO under stronger condition.
2. *Matrix ℓ_1-norm guarantees* for CGLASSO-1 and CGLASSO-3 under weaker condition.

Theoretical guarantees

Bound on estimation error $\|\hat{\Theta}(\omega_j) - \Theta^*(\omega_j)\|_\infty$.

1. *Entry-wise max norm guarantees* for vanilla CGLASSO under stronger condition.
2. *Matrix ℓ_1-norm guarantees* for CGLASSO-1 and CGLASSO-3 under weaker condition.

Key technical pieces:

- Single deviation bound on $\|\hat{f}(\omega_j) - f(\omega_j)\|_\infty$ (Sun et al., 2018).
- **Vanilla CGLASSO**: α-incoherence condition on Hessian of the Whittle log-likelihood (Ravikumar et al., 2011).
- **CGLASSO-3**: consistency of $\hat{\tau}_k^2$ (van de Geer et al., 2014).

CGLASSO consistency

- $\alpha \in (0, 1]$: incoherence parameter, $C_\alpha = 1 + 8/\alpha$.
- Stability parameter: $|||f||| = \operatorname{ess\,sup}_{\omega \in [-\pi, \pi]} \|f(\omega)\|_2$.
- Model parameters: $\kappa_f, \kappa_{\Theta^*}$, temporal dependence parameters: L_n, Ω_n.
- s: Number of edges, d: max-row degree.
- Threshold: $\Delta = |||f||| \sqrt{C \log p/m} + \frac{m}{n}\Omega_n + L_n$.

CGLASSO consistency

- $\alpha \in (0, 1]$: incoherence parameter, $C_\alpha = 1 + 8/\alpha$.
- Stability parameter: $|||f||| = \operatorname{ess\,sup}_{\omega\in[-\pi,\pi]} \|f(\omega)\|_2$.
- Model parameters: $\kappa_f, \kappa_{\Theta^*}$, temporal dependence parameters: L_n, Ω_n.
- s: Number of edges, d: max-row degree.
- Threshold: $\Delta = |||f|||\sqrt{C \log p/m} + \frac{m}{n}\Omega_n + L_n$.

Theorem (Simplified)

Let X_t be a p-dim linear Gaussian time series satisfying α-incoherence condition. Under certain regularity condition with $n/\Omega_n \gtrsim m \gtrsim \Omega_n |||f|||^2 \log p$, $\lambda = (8/\alpha)\Delta$, and $\Delta d \le 1/(6\kappa^2_{\Theta^*}\kappa^3_f C_\alpha)$, with high probability,

$$\left\|\hat{\Theta} - \Theta^*\right\|_\infty \le 2\kappa_{\Theta^*} C_\alpha \Delta,$$

$$\{|\Theta^*_{k,\ell}| > 2\kappa_{\Theta^*} C_\alpha \Delta\} \subset E(\hat{\Theta}) \subset E(\Theta^*).$$

CGLASSO consistency

- $\alpha \in (0, 1]$: incoherence parameter, $C_\alpha = 1 + 8/\alpha$.
- Stability parameter: $|||f||| = \operatorname{ess\,sup}_{\omega \in [-\pi, \pi]} \|f(\omega)\|_2$.
- Model parameters: $\kappa_f, \kappa_{\Theta^*}$, temporal dependence parameters: L_n, Ω_n.
- s: Number of edges, d: max-row degree.
- Threshold: $\Delta = |||f||| \sqrt{C \log p/m} + \frac{m}{n}\Omega_n + L_n$.

Theorem (Simplified)

Let X_t be a p-dim linear Gaussian time series satisfying α-incoherence condition. Under certain regularity condition with $n/\Omega_n \gtrsim m \gtrsim \Omega_n |||f|||^2 \log p$, $\lambda = (8/\alpha)\Delta$, and $\Delta d \leq 1/(6\kappa_{\Theta^*}^2 \kappa_f^3 C_\alpha)$, with high probability,

$$\left\|\hat{\Theta} - \Theta^*\right\|_\infty \leq 2\kappa_{\Theta^*} C_\alpha \Delta,$$

$$\{|\Theta^*_{k,\ell}| > 2\kappa_{\Theta^*} C_\alpha \Delta\} \subset E(\hat{\Theta}) \subset E(\Theta^*).$$

Remark. If $m \asymp n^\xi$ for $\xi > 0$, and $\|\Gamma(h)\| \leq \sigma\rho^{|h|}$, $\Delta \asymp \sqrt{\log p/m}$ and $m \gtrsim d^2 \log p$.

Consistency of the adaptive variant

- **Assume bounded spectrum**: $\max\{|||f|||, |||\Theta|||\} \leq M$.
- Nodewise regression penalty for CGLASSO-3: $\lambda_j \asymp M\sqrt{\log p/m}$.

Consistency of the adaptive variant

- **Assume bounded spectrum**: $\max\{|||f|||, |||\Theta|||\} \leq M$.
- Nodewise regression penalty for CGLASSO-3: $\lambda_j \asymp M\sqrt{\log p/m}$.

Theorem (Simplified)

Let X_t be a p-dim linear Gaussian time series with bounded spectrum. Under certain regularity condition with $n/\Omega_n \gtrsim m \gtrsim M^6 d\log p$, $s\lambda \lesssim 1/M^5$, with high probability,

CGLASSO-1: *For* $\lambda \asymp M^4\sqrt{\log p/m}$, $\left|\left|\left|\hat{\Theta}_D - \Theta^*\right|\right|\right|_1 \leq C_M s\lambda$,

CGLASSO-3: *For* $\lambda \asymp M^4\sqrt{d\log p/m}$, $\left|\left|\left|\hat{\Theta}_D - \Theta^*\right|\right|\right|_1 \leq C'_M(p+s)\lambda$.

Consistency of the adaptive variant

- **Assume bounded spectrum**: $\max\{|||f|||, |||\Theta|||\} \le M$.
- Nodewise regression penalty for CGLASSO-3: $\lambda_j \asymp M\sqrt{\log p/m}$.

Theorem (Simplified)

Let X_t be a p-dim linear Gaussian time series with bounded spectrum. Under certain regularity condition with $n/\Omega_n \gtrsim m \gtrsim M^6 d\log p$, $s\lambda \lesssim 1/M^5$, with high probability,

$$\text{CGLASSO-1:} \quad \textit{For } \lambda \asymp M^4\sqrt{\log p/m}, \quad \left|\left|\left|\hat{\Theta}_D - \Theta^*\right|\right|\right|_1 \le C_M s\lambda,$$

$$\text{CGLASSO-3:} \quad \textit{For } \lambda \asymp M^4\sqrt{d\log p/m}, \quad \left|\left|\left|\hat{\Theta}_D - \Theta^*\right|\right|\right|_1 \le C'_M (p+s)\lambda.$$

Remark. $s\lambda \lesssim 1/M^5$: stronger condition on the sparsity.

$$\text{CGLASSO-1} : m \gtrsim s^2 \log p, \quad \text{CGLASSO-3} : m \gtrsim s^2 d\log p.$$

Main takeaways

Our contribution:

- Fast pathwise coordinate descent algorithm for complex variable lasso.
- Frequency-domain graphical lasso using CLASSO.
- Adaptive CGLASSO.

Main takeaways

Our contribution:

- Fast pathwise coordinate descent algorithm for complex variable lasso.
- Frequency-domain graphical lasso using CLASSO.
- Adaptive CGLASSO.

Future directions:

- Inference of CGLASSO.
- Bandwidth selection.
- Aggregation across frequencies + recovery of the true graph for stationary time series.
- Non-stationary GGMs (Basu and Subba Rao, 2023).

Main takeaways

Our contribution:

- Fast pathwise coordinate descent algorithm for complex variable lasso.
- Frequency-domain graphical lasso using CLASSO.
- Adaptive CGLASSO.

Future directions:

- Inference of CGLASSO.
- Bandwidth selection.
- Aggregation across frequencies + recovery of the true graph for stationary time series.
- Non-stationary GGMs (Basu and Subba Rao, 2023).

Paper: `https://arxiv.org/abs/2401.11128`
R Package `cxreg`: `https://github.com/navonildeb/cxreg`

Thank you!

Navonil Deb

E-mail: nd329@cornell.edu

References

Baek, C., Düker, M.-C., and Pipiras, V. (2021). Thresholding and graphical local whittle estimation. *arXiv preprint arXiv:2105.13342*.

Basu, S. and Subba Rao, S. (2023). Graphical models for nonstationary time series. *The Annals of Statistics*, 51(4):1453–1483.

Böhm, H. and von Sachs, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, 100(5):913–935.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, 51(2):157–172.

Fiecas, M., Leng, C., Liu, W., and Yu, Y. (2019). Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics*, 13(2):4079–4101.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Janková, J. and van de Geer, S. (2018). Inference in high-dimensional graphical models. In *Handbook of graphical models*, pages 325–350. CRC Press.

Krampe, J. and Paparoditis, E. (2022). Frequency domain statistical inference for high-dimensional time series. *arXiv preprint arXiv:2206.02250*.

Maleki, A., Anitori, L., Yang, Z., and Baraniuk, R. G. (2013). Asymptotic analysis of complex lasso via complex approximate message passing (camp). *IEEE Transactions on Information Theory*, 59(7):4290–4308.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, 34(3):1436–1462.

Ombao, H. and Pinto, M. (2022). Spectral dependence. *Econometrics and Statistics*.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

Scheinberg, K., Ma, S., and Goldfarb, D. (2010). Sparse inverse covariance selection via alternating linearization methods. *Advances in neural information processing systems*, 23.

Sun, Y., Li, Y., Kuceyeski, A., and Basu, S. (2018). Large spectral density matrix estimation by thresholding. *arXiv preprint arXiv:1812.00532*.

Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., and Tibshirani, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(2):245–266.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166 – 1202.