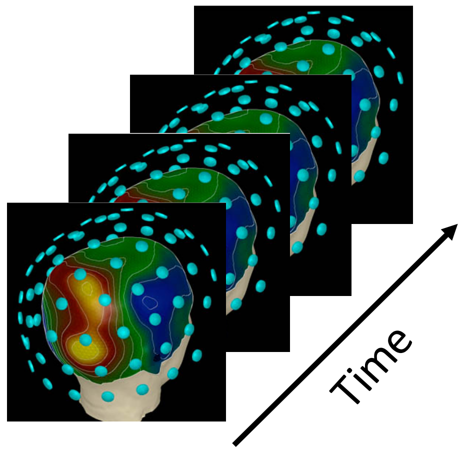
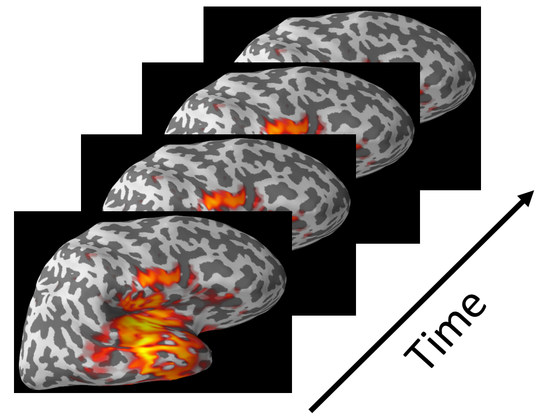
Sparse plus low-rank graphical models of time series

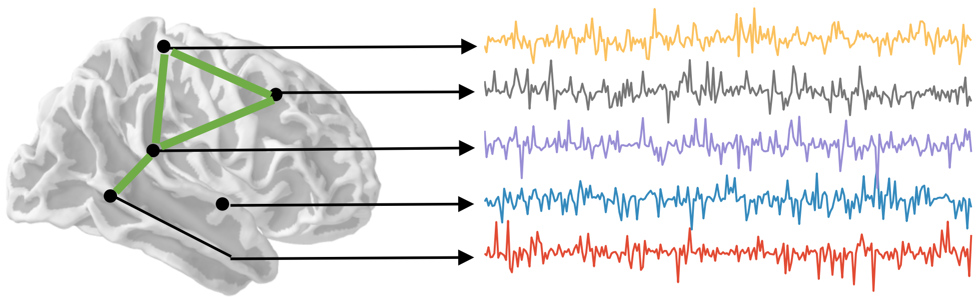
Presented by Rahul Nadkarni
Joint work with Nicholas J. Foti, Adrian KC Lee, and Emily B. Fox
University of Washington
August 14th, 2016

Brain Interactions from MEG

Magnetoencephalography (MEG) captures weak magnetic field.







Goal: Infer functional connectivity

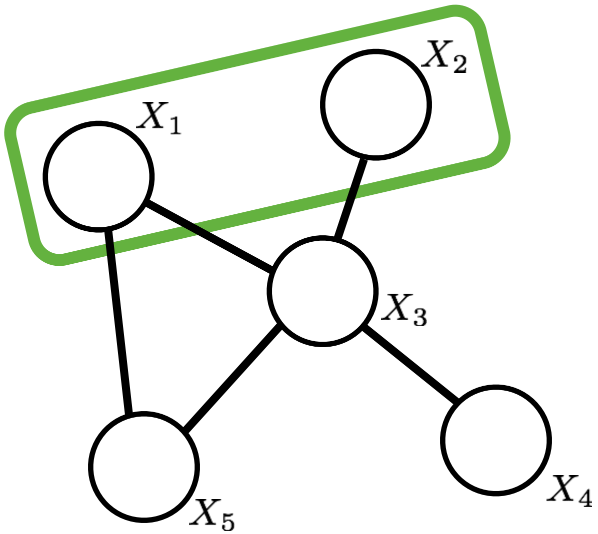
Graphical Models

- Graph G=(V, E) encodes conditional independence statements.

nodes edges

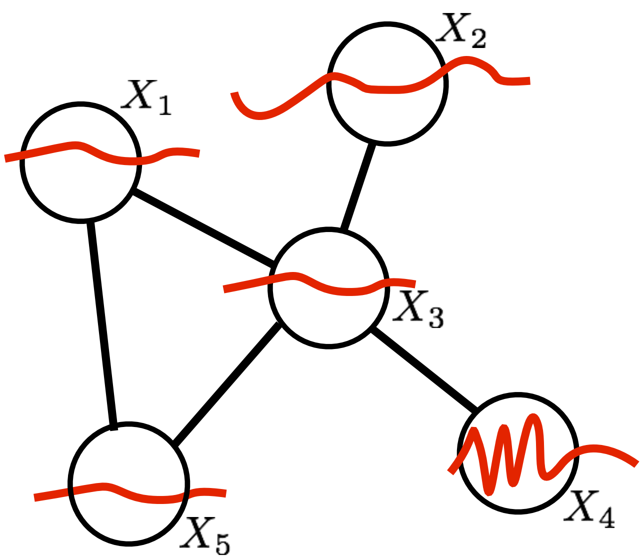
No edge (i, j) $\longleftrightarrow$ X_i , X_j conditionally independent given rest of variables.

$$X_1 \perp\!\!\!\perp X_2 | X_3, X_4, X_5$$



Graphical Models of Time Series

No edge (i,j) $\Rightarrow$ **time series** X_i , X_j conditionally independent given *entire trajectories* of other series.

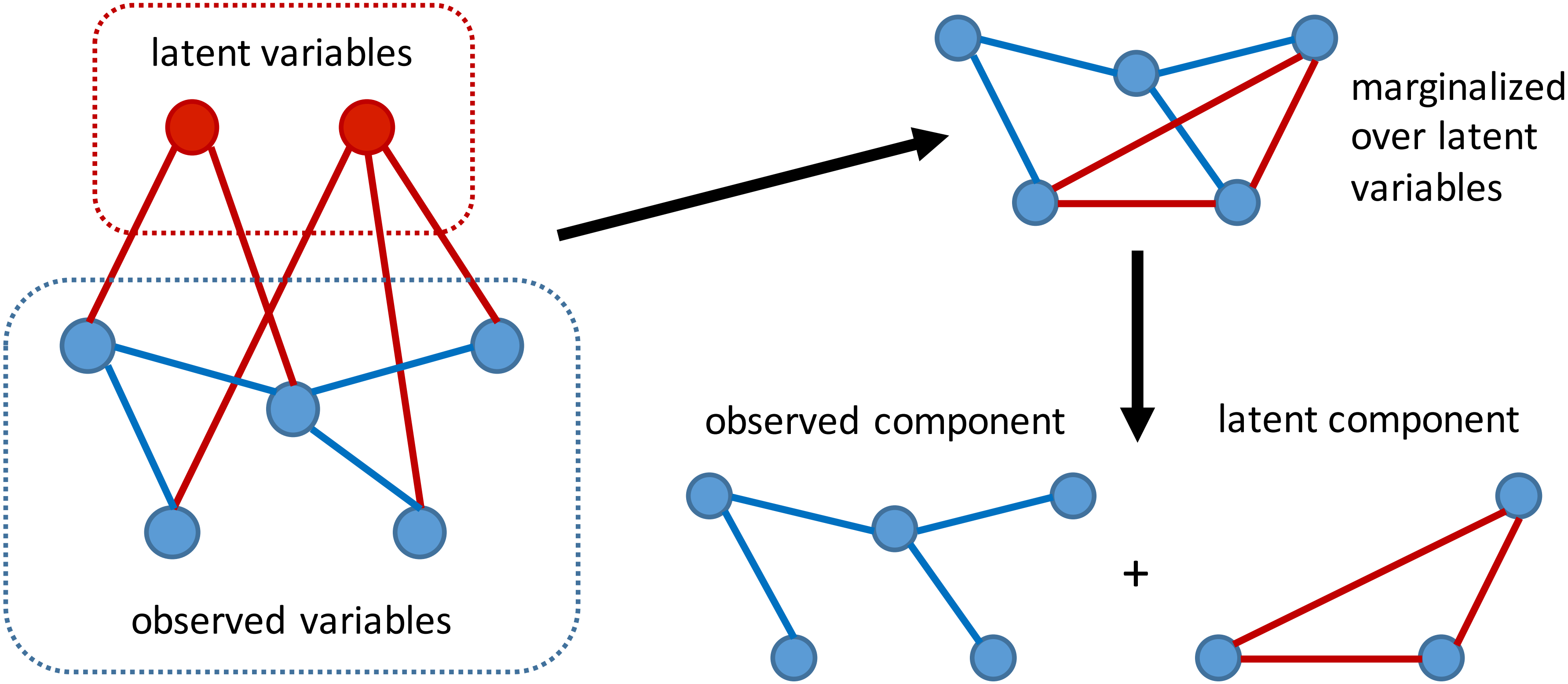


- Accounts for interactions at *all lags*.
- Removes linear effects of other series.

Natural property for **functional connectivity**

Examples of existing work:
Bach et al. 2004, Songsiri & Vandenberghe 2010, Jung et al. 2015, Tank et al. 2015

Latent structure



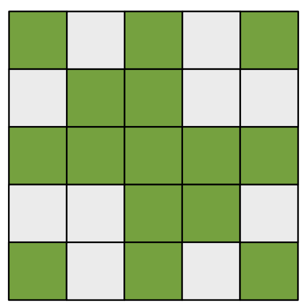
Examples of existing work:
Chandrasekaran et al. 2012, Jalali & Sanghavi 2012, Liégois et al. 2015

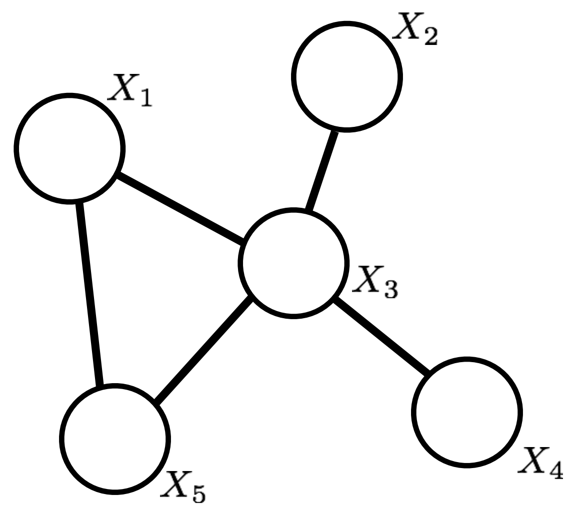
Encoding graph structure

Gaussian random vectors

- For Gaussian random vector $X \sim \mathcal{N}(0, \Sigma)$

Conditional independence encoded in the precision matrix.

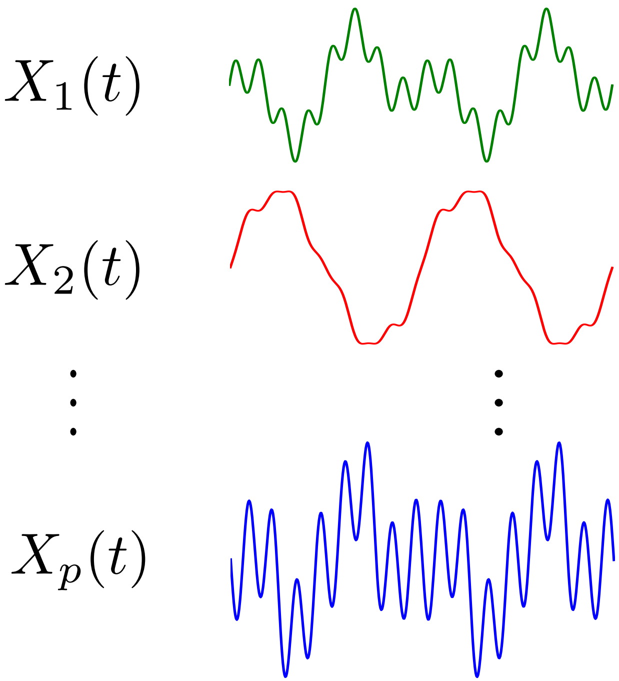
$\Sigma^{-1} =$ 



$$(\Sigma^{-1})_{ij} = 0$$
$$\Updownarrow$$

X_i , X_j conditionally independent given rest of variables.

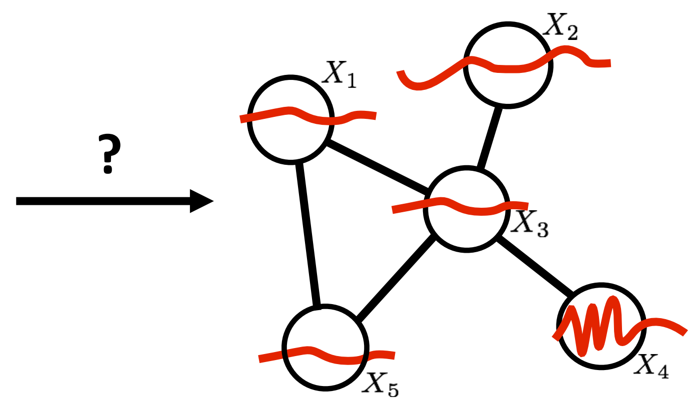
Gaussian stationary processes



lagged covariance:

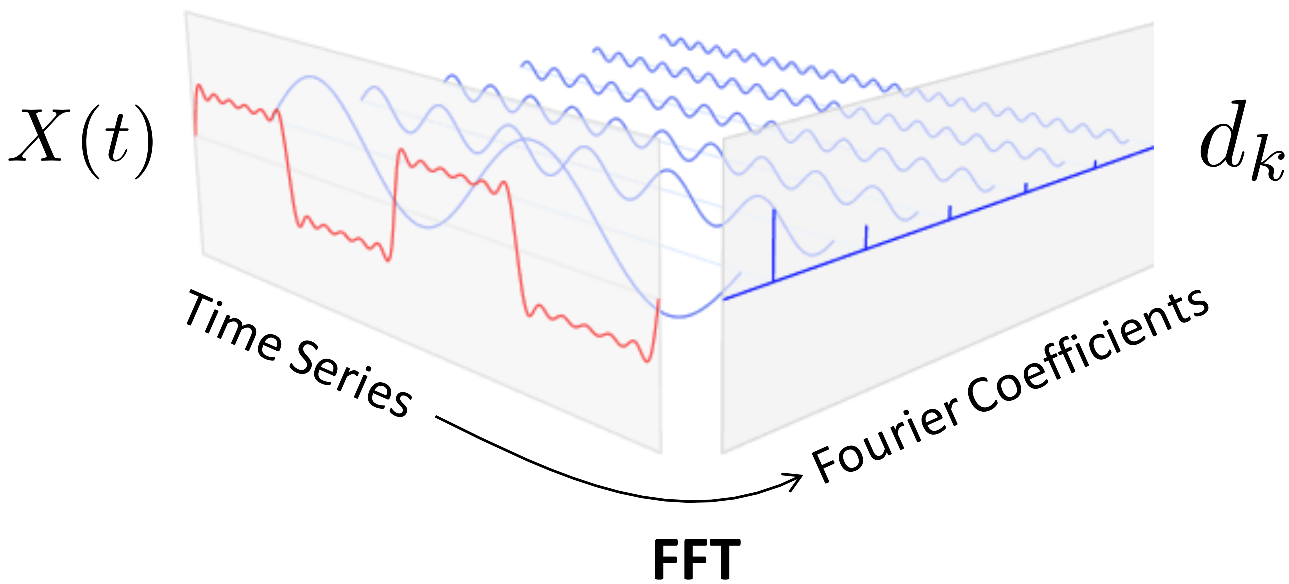
$$\Gamma(h) = \mathrm{Cov}(X(t), X(t+h))$$

$$\Gamma(h)$$



How is conditional independence encoded?

Model in the Frequency Domain



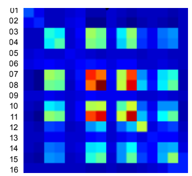
Spectral density matrix

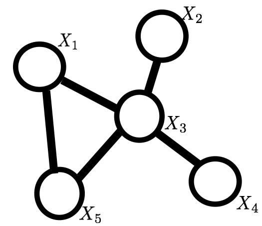
Lagged covariance matrix

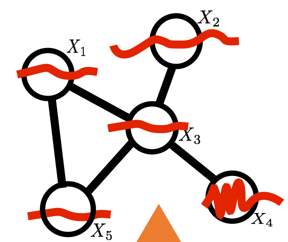
$$\Gamma(h) = \mathrm{Cov}(X(t), X(t+h)) \longrightarrow S(\lambda) = \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-i\lambda h}$$

Encoding structure in frequency domain

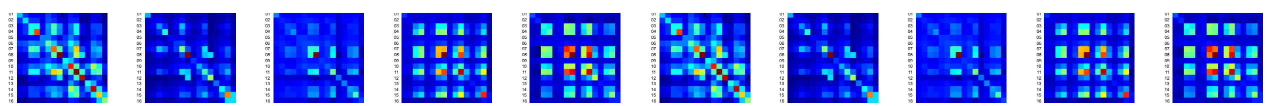
For **Gaussian** i.i.d. random variables,

$\Sigma^{-1} =$ 





(*Dahlhaus, 2000*) For **Gaussian** stationary time series,
$S(\lambda)^{-1}$:



$\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_k \quad \cdots \quad \lambda_T$

complex inverse **spectral density** matrices

Learning structure from data

Penalized likelihood expression

inverse covariance matrix

$$-\log\left(\text{Likelihood}(\Psi)\right) \quad + \quad \lambda \cdot \text{Penalty}(\Psi)$$

sample covariance matrix

$$-\log\det\Psi + \text{tr}\left\{\hat{S}\Psi\right\} \qquad \sum_{i<j}|\Psi_{ij}|$$

negative log-likelihood of Gaussian

sparsity-inducing penalty

Graphical LASSO (Friedman et al. 2007)

solved with: many existing algorithms

What's our likelihood in the frequency-domain case?

Likelihood in the Frequency Domain

Time Domain Likelihood

$$p(X(1),\ldots,X(T)|[\Gamma(h)]_{h=0}^{T-1})$$

$\longrightarrow$

Frequency Domain Likelihood

$$p(\underbrace{d_0,\ldots,d_{T-1}}_{\text{Fourier coefficients}}|\{S(\lambda_k)\}_{k=0}^{T-1})$$

Fourier coefficients are asymptotically independent, complex Normal random vectors (Brillinger, 1981)

Whittle Approximation

$$S_k \equiv S(\lambda_k)$$

$$p(d_1,\ldots,d_T|\{S(\lambda_k)\}_{k=0}^{T-1}) \approx \prod_{k=0}^{T-1} \frac{1}{\pi^p|S_k|} e^{-d_k^* S_k^{-1} d_k}$$

Penalized likelihood expression in frequency domain

inverse spectral density matrix

$$-\log\left(\prod_{k=0}^{T-1}\frac{1}{\pi^p|S_k|}e^{-d_k^* S_k^{-1} d_k}\right) \quad + \quad \lambda\cdot\mathrm{Penalty}(\Psi)$$

sample spectral density matrix

$$\sum_{k=0}^{T-1}\Big(-\log\det\Psi[k]+\mathrm{tr}\left\{\hat{S}[k]\Psi[k]\right\}\Big) \qquad \sum_{i<j}\sqrt{\sum_{k=0}^{T-1}|\Psi[k]_{ij}|^2}$$

Group LASSO penalty

Spectral graphical LASSO (Jung et al. 2015)

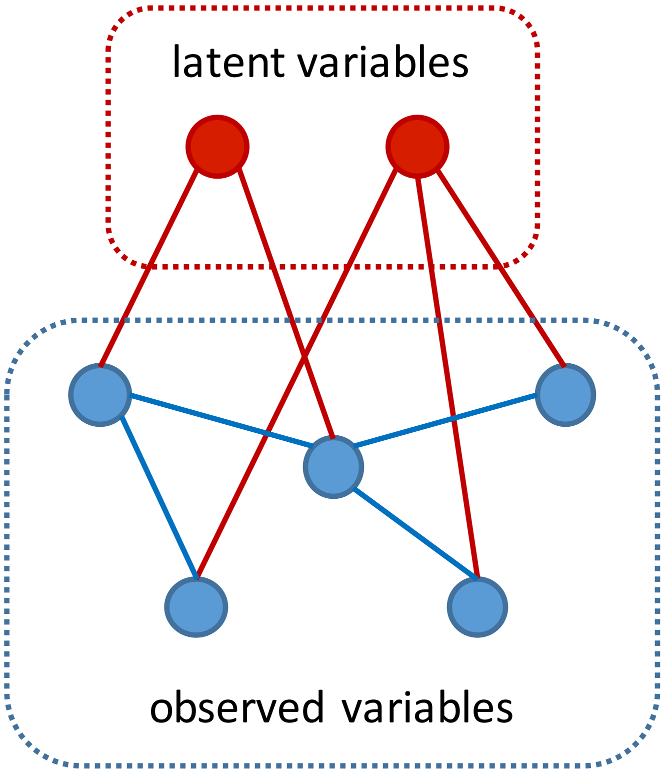
solved with: ADMM (Jung et al. 2015)

Incorporating latent processes

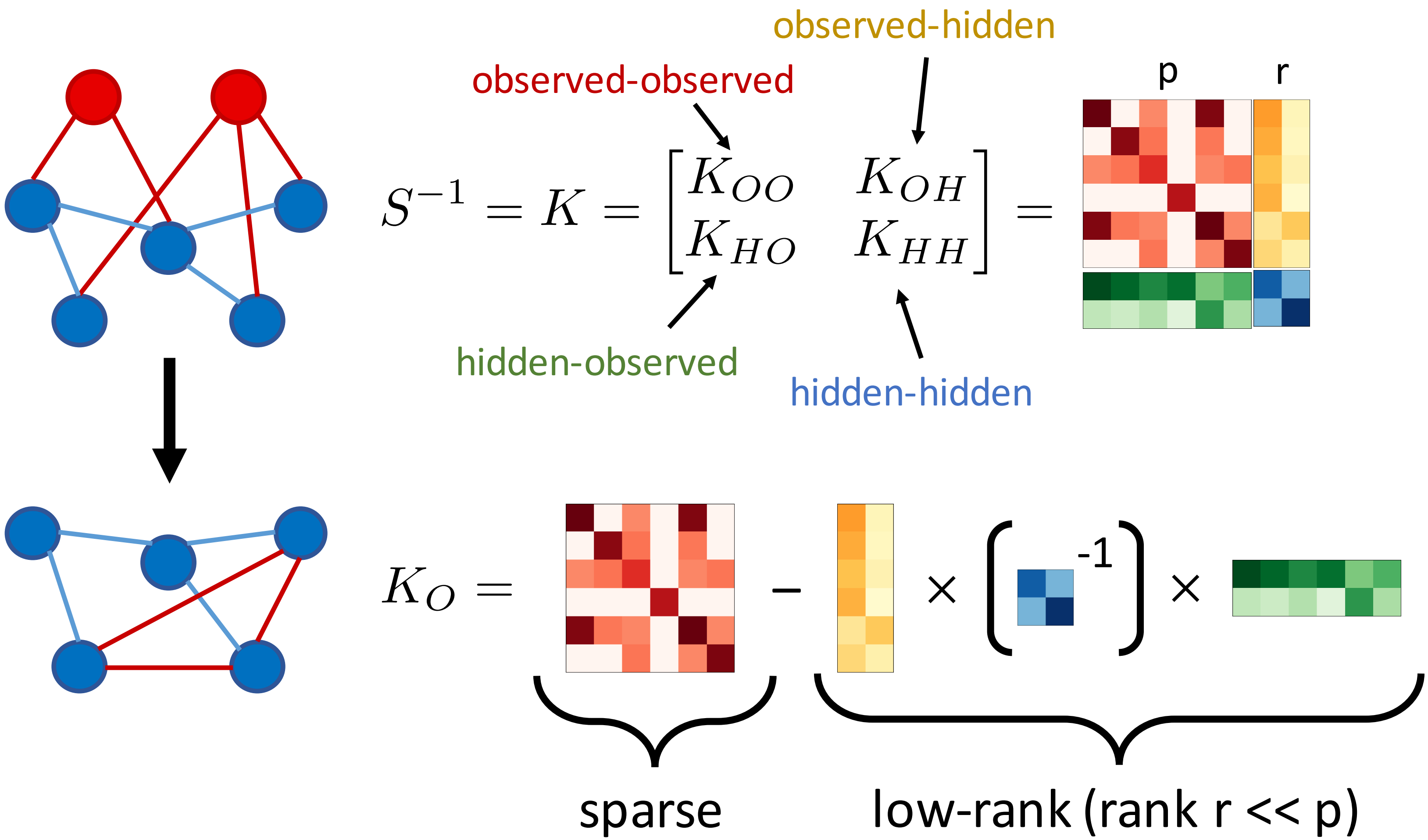
Latent structure in MEG

- MEG recordings affected by neural activity unrelated to task
- Mapping from recordings to brain activity introduces "point spread"

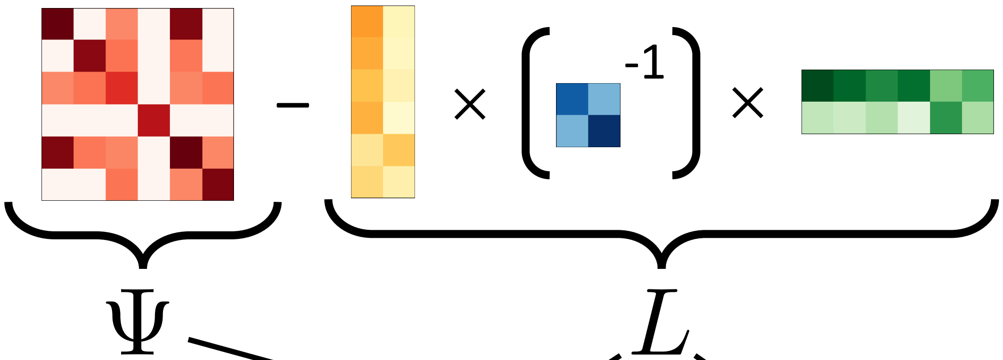
These issues can be addressed by adding a latent component to the model



Sparse plus low-rank decomposition



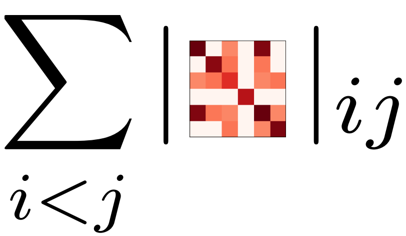
Sparse plus low-rank penalized likelihood



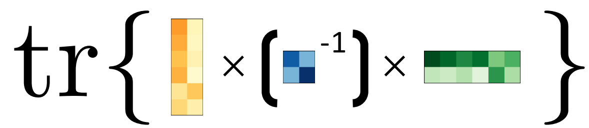
negative log-likelihood:

$$-\log\det(\Psi - L) + \mathrm{tr}\left\{\hat{S}(\Psi - L)\right\}$$

sparse penalty:



low-rank penalty:



Latent variable GLASSO (Chandrasekaran et al. 2012)

solved with ADMM (Ma et al. 2013)

Latent variable spectral GLASSO

negative log-likelihood:

$$\sum_{k=0}^{T-1} \left(-\log\det(\Psi[k] - L[k]) + \mathrm{tr}\left\{ \hat{S}[k](\Psi[k] - L[k]) \right\} \right)$$

Whittle approximation

sparse penalty:

$$\sum_{i<j} \sqrt{\sum_{k=0}^{T-1} |\Psi[k]_{ij}|^2}$$

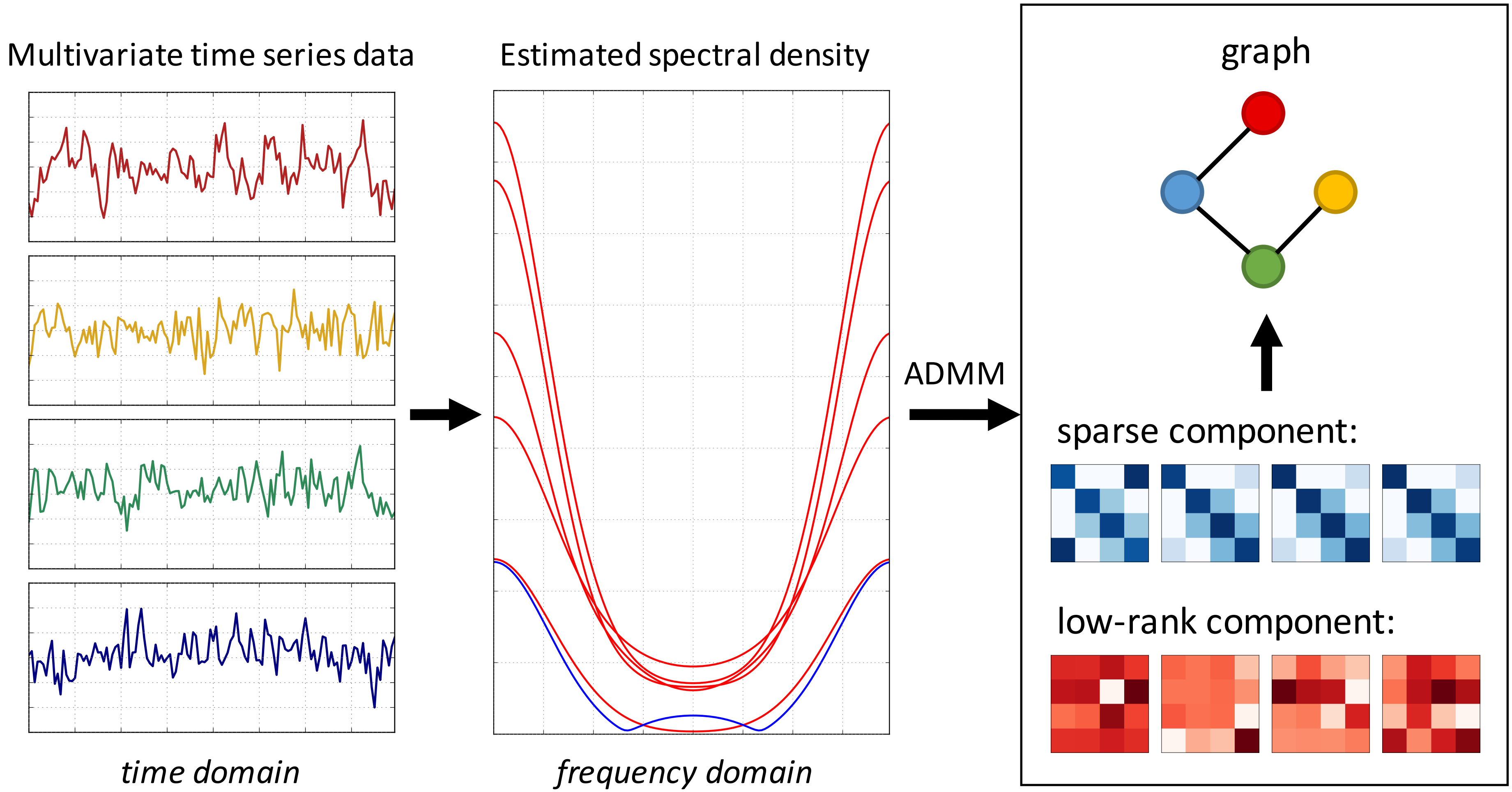
Group LASSO penalty

low-rank penalty:

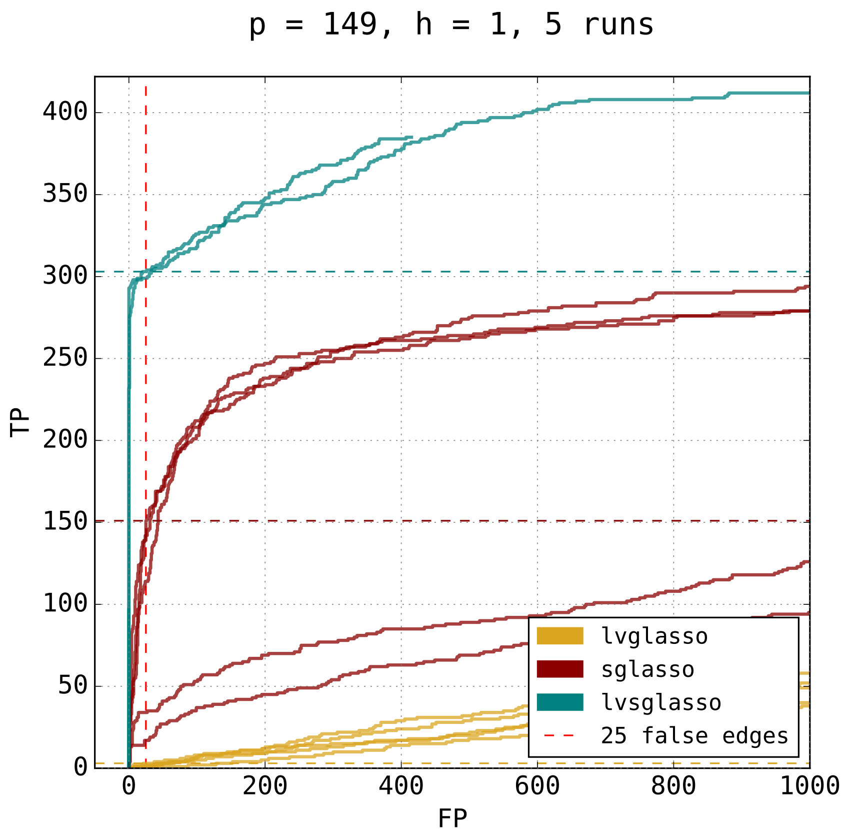
$$\sum_{k=0}^{T-1} \mathrm{tr}\left\{ L[k] \right\}$$

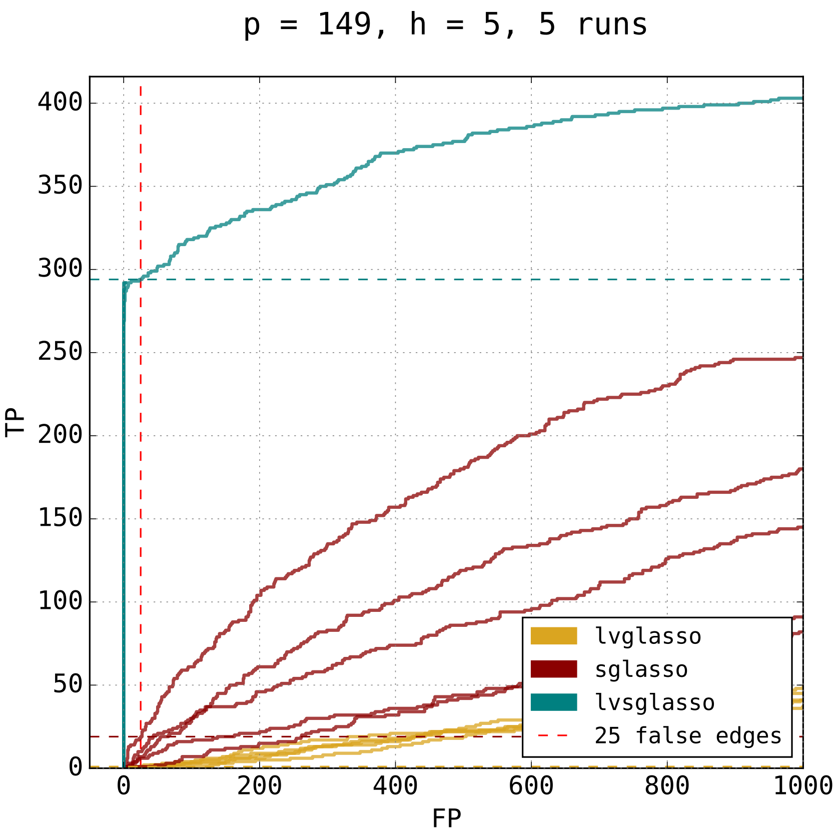
Used ADMM to solve this convex formulation

Analysis pipeline



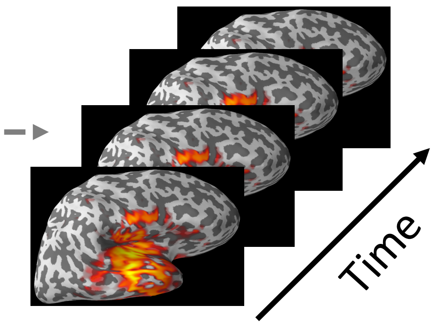
Synthetic data results



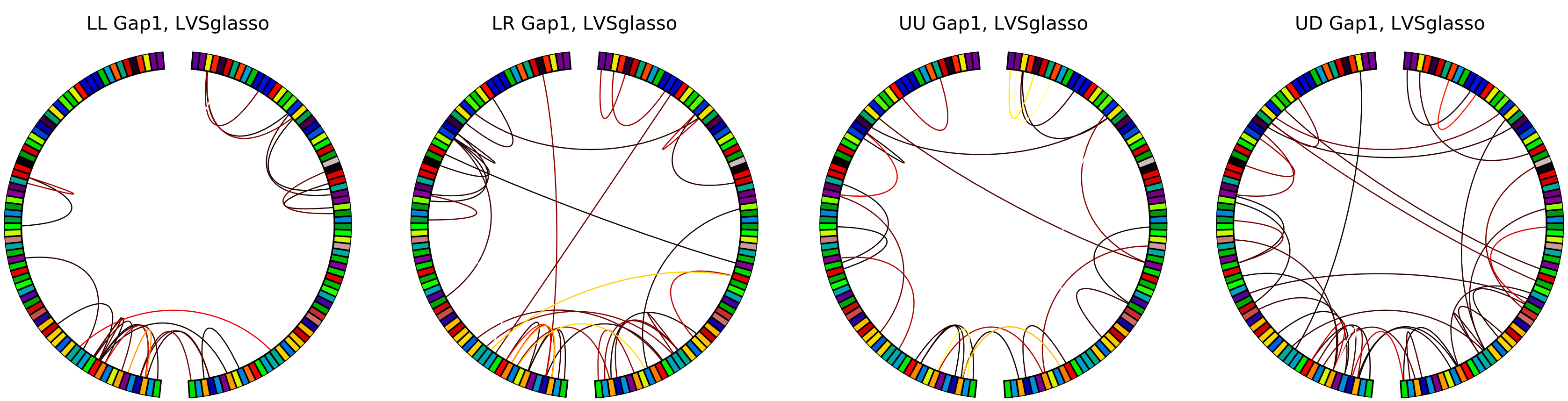


MEG Auditory Attention Analysis

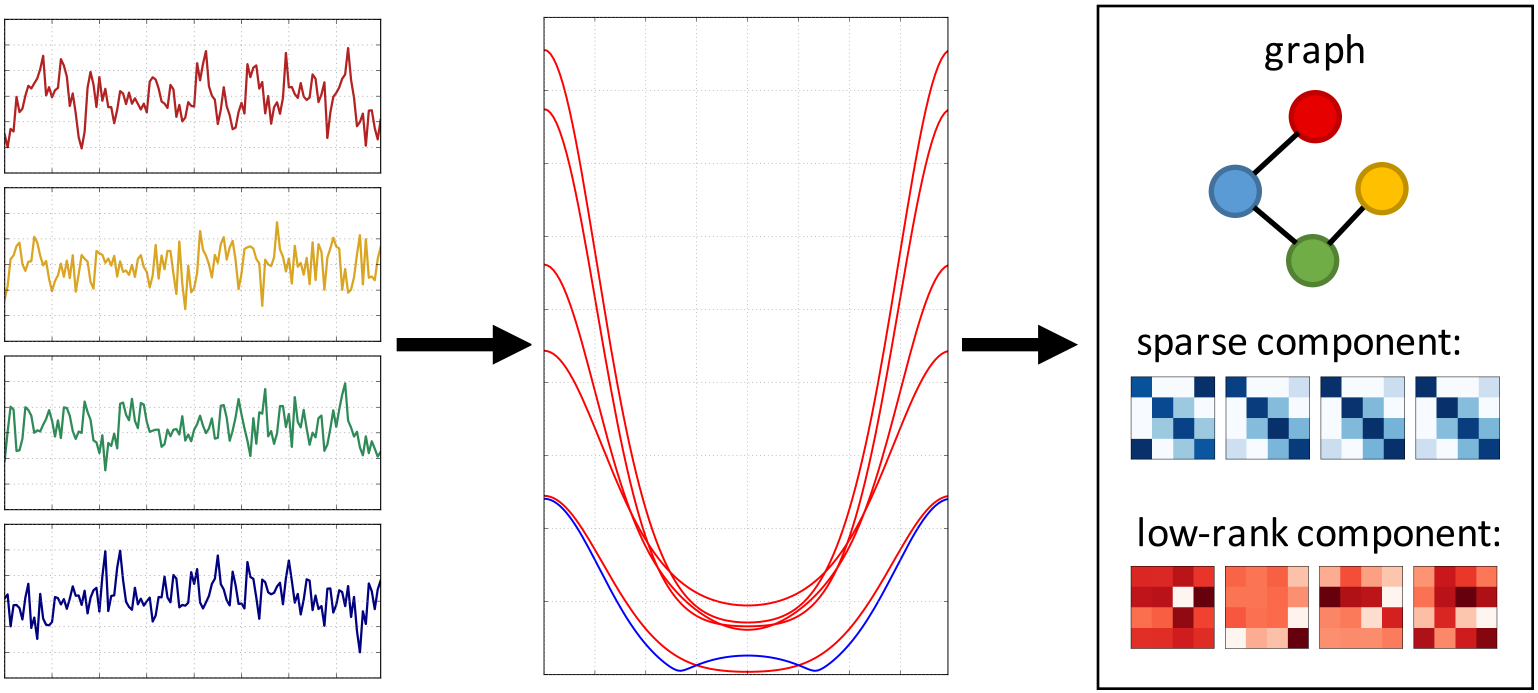
Maintain or Switch attention
(Left/Right, High/Low pitch)



- 16 subjects, 10-50 trials each.
- Each trial results in a 149-dimensional time series.



Summary



- Frequency domain for conditional independence structure and likelihood
- Modeling latent component gives sparser, more interpretable graphs
- Latent variable, spectral models are important in neuroscience