Reading group on brain networks: From imaging data to connectivity graphs

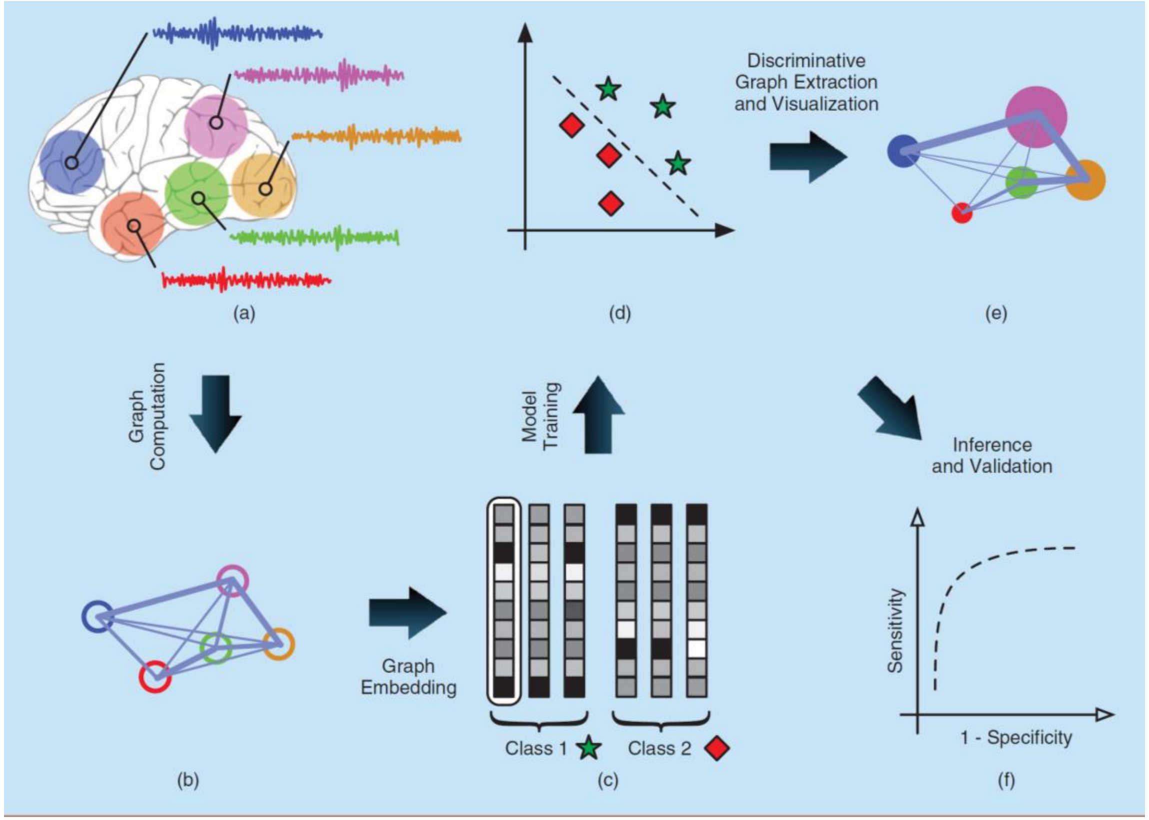
Fateme Sheikholeslami

Dept. of ECE and DTC, University of Minnesota

J. Richardi, S. Archad, H. Bunke, and D. Van De Ville, “Machine learning with graphs”, SP Magazine, 2012.

May 21st, 2015

Recap



Question: How to form graph vertices and edges?

Mathematical definition

- General graph representation

$$g = (V_g, E_g, \alpha_g, \beta_g)$$

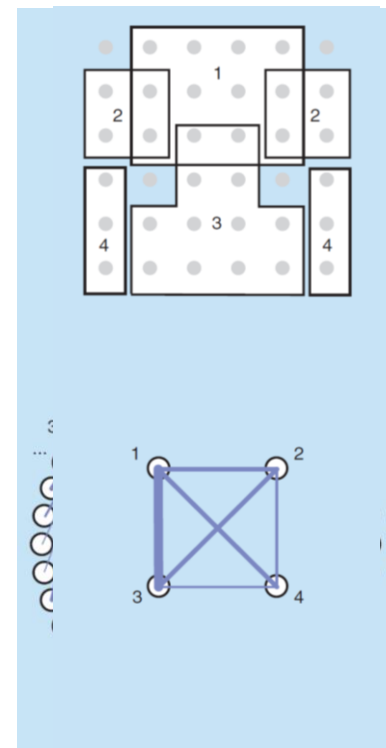
- Graph vertices V_g
- Graph edges E_g
- Vertex labeling function $\alpha_g : V_g \to L_v$
  - Notion of graph-theoretical attributes (e.g. vertex centrality)
- Edge labeling function $\beta_g : E_g \to L_E$
  - Notion of statistical dependence between nodes

Graph Vertices

- Voxel per voxel, a.k.a. seed based
  - Spatially contiguous subsets

- Anatomy-driven: Atlas-based
  - Spatially contiguous subsets
  - Typically in the low hundreds of vertices

- Data-driven
  - Spatial ICA, PCA, clustering
  - Example: ICA yields around 20 ents
  - May have contiguous subsets



Principle Component analysis

- Data $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_V]$

$$\mathrm{cov}(\mathbf{X}) = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$$

$$[\mathbf{u}_1, \ldots, \mathbf{u}_k] = \mathrm{evecs}(\mathbf{X}\mathbf{X}^\top)$$

$$\mathbf{u}_1(t) = \sum_{v=1}^{V} \alpha_i \mathbf{x}_i(t)$$

- A means of dimensionality reduction
- A means of measuring functional connectivity
- Works best for Gaussian signals

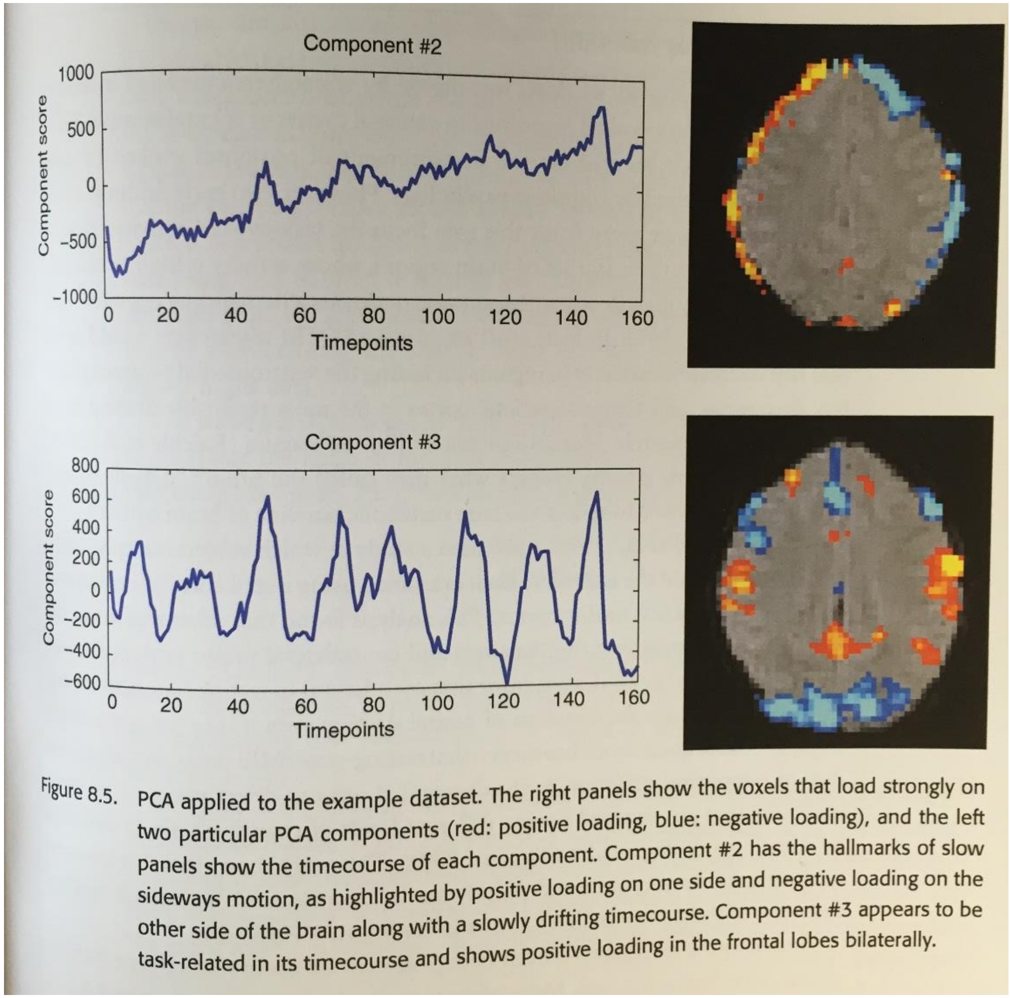
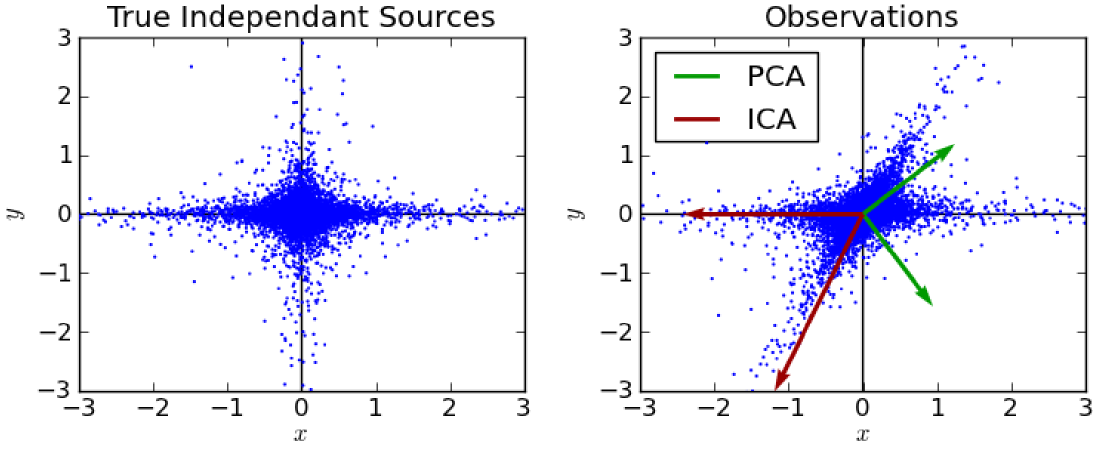


Figure 8.5. PCA applied to the example dataset. The right panels show the voxels that load strongly on two particular PCA components (red: positive loading, blue: negative loading), and the left panels show the timecourse of each component. Component #2 has the hallmarks of slow sideways motion, as highlighted by positive loading on one side and negative loading on the other side of the brain along with a slowly drifting timecourse. Component #3 appears to be task-related in its timecourse and shows positive loading in the frontal lobes bilaterally.

R. Poldrack et al, “Handbood of Functional MRI”, Cambridge University Press, 2011

PCA vs. ICA



- Principle Component Analysis
  - Captures orthogonal directions with the highest variance
  - Assumes Guassianity on data
- Independent Component Analysis
  - Relies on non-Gaussianity assumption
  - Minimizes the mutual information of the output
  - Very similar to blind source separation problems

Blind source separation

- Objective

  Separating a set of unknown source signals from mixed observations

- System model

$$x_1(t) = a_{11}s_a + a_{12}s_2$$
$$x_2(t) = a_{21}s_a + a_{22}s_2 \quad \Rightarrow \quad \mathbf{x} = \mathbf{A}\mathbf{s} \quad \overset{\text{BSS}}{\Rightarrow} \quad \mathbf{s} = \mathbf{W}\mathbf{x}$$

- Independent component analysis
  - Uses higher order statistics rather than just covariance
  - Builds on statistical independence of the sources
  - Find $\mathbf{w}_j$ such that non-gaussianity of $\mathbf{w}_j^\top \mathbf{x}$ is maximized (e.g. negentropy, kurtosis, etc.)
  - Ambiguity in (1) scaling and (2) source ordering

Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13.4 (2000): 411-430.

Example: FastICA

1. Center **x** (remove the mean from **x**)
2. Whiten **x** (uncorrelated the components)
3. for i = 1 to n

    w = random vector;
    orthogonalize initial vector w in terms of the previous components;
    normalize w;
    while (w not converged)
        w = approximation of negentropy of $w^T\mathbf{x}$
        orthogonalize w in terms of the previous components;
        normalize w;
    end while
    W (i,:) = w;

   end for
4. s = W*whitenedx; return s;

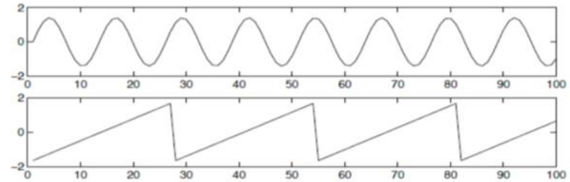


Figure 1: The original signals.

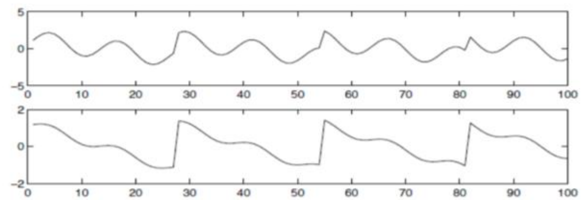


Figure 2: The observed mixtures of the source signals in Fig. 1.

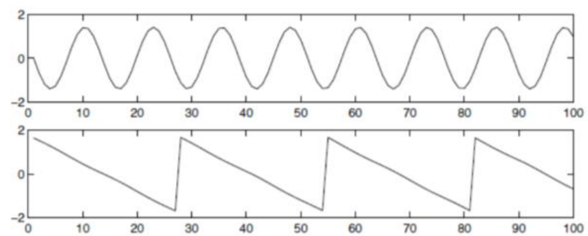


Figure 3: The estimates of the original source signals, estimated using only the observed signals in Fig. 2. The original signals were very accurately estimated, up to multiplicative signs.

Independent component analysis (Cnt'd)

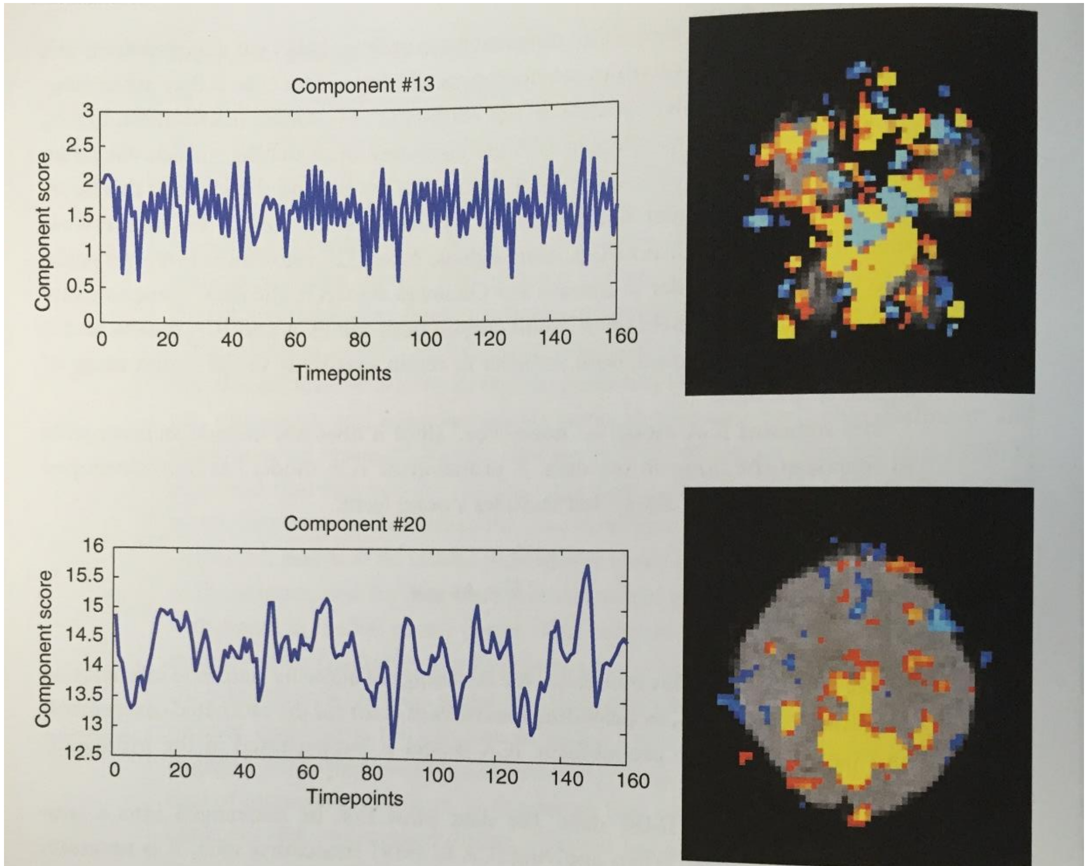


Figure 8.6. ICA applied to the example dataset. The top panels show data from a component that appears to represent high-frequency oscillations, most likely due to cardiac pulsation. This component loaded most heavily in the most inferior slices covering the brain stem, as shown in the slice image. The bottom panels show a component that has a potentially task-relevant waveform and show heaviest loading in the medial parietal and intraparietal regions. ICA was performed using using the FastICA toolbox for MATLAB.

R. Poldrack *et al.*, "Handbood of Functional MRI", Cambridge University Press, 2011

Vertex time series

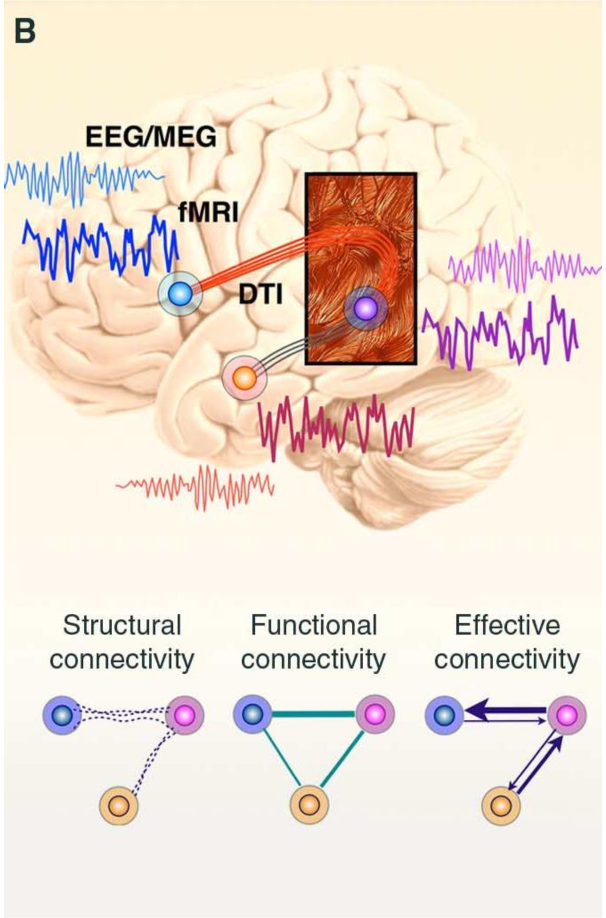
- Seed-based vertices
  - Spatial smoothing to increase signal-to-noise ratio

- Atlas-based vertices
  - Temporal mean of the times series of voxels in a vertex
  - First eigenvariate as the vertex representative

- Data-based vertices

Edge label assignments

- Inferring statistical dependencies between brain regions



- Functional connectivity

  Presence or absence of interaction between each pair of regions

- Effective connectivity

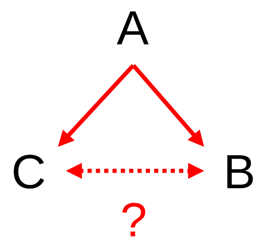
  The directionality (causality) of the existing interconnections

Park, Hae-Jeong, and Karl Friston. "Structural and functional brain networks: from connections to cognition." Science 342.6158 (2013): 1238411.

Correlation and partial correlation

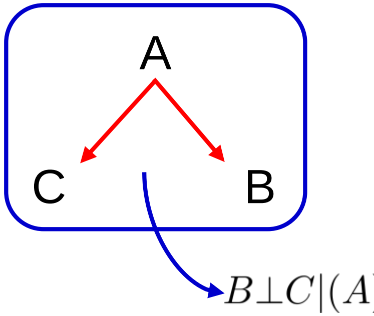
- Pearson product moment correlation

$$\rho_{i,j} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

  - What if 

?
  - Correlation $\not\Longrightarrow$ Causality!

- Partial correlation: Correlation of a pair of nodes given the rest of the network



Correlation and partial correlation (Cnt'd)

- Edges $(\mathbf{x}_i, \mathbf{x}_j) \notin E \Longleftrightarrow (\mathbf{x}_i, \mathbf{x}_j)$ conditionally uncorrelated given $\{\mathbf{x}_k : k \neq i, k \neq j\}$

- Partial correlation

$$\rho_{i,j|(V\setminus\{i,j\})} = \frac{\sigma_{ij|(V\setminus\{i,j\})}}{\sqrt{\sigma_{ii|(V\setminus\{i,j\})}\sigma_{ij|(V\setminus\{i,j\})}}}$$

- If $\operatorname{cov}\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ then $\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$\Longrightarrow$ $\boxed{\text{Conditional independence graphs}}$

Kolaczyk, Eric D. *Statistical analysis of network data: methods and models*. Springer Science & Business Media, 2009.

Correlation and partial correlation (Cnt'd)

❖ Gaussian multivariate r.v.s

- Edges $(\mathbf{x}_i, \mathbf{x}_j) \notin E \Longleftrightarrow (\mathbf{x}_i, \mathbf{x}_j)$ conditionally independent given $\{\mathbf{x}_k : k \neq i, k \neq j\}$

$\Longrightarrow$ $\boxed{\text{Gaussian graphical models}}$

- <u>Remark</u>

$$\rho_{i,j|(V\backslash\{i,j\})} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\,\omega_{jj}}} \quad \text{where} \quad \omega_{ij} = [\Omega]_{i,j} = [\Sigma^{-1}]_{i,j}$$

Concentration or *precision* matrix

$\Longrightarrow$ $\boxed{\text{Concentration graphs}}$

Inverse covariance

Partial correlation for the entire network $\Longrightarrow$ Inverse covariance matrix estimation

Pair-wise modeling

Global network modeling

Pair-wise full corr.

Partial corr.

Inv. Cov.

SEM

DCM

- Maximum likelihood estimation of Σ^{-1}

$$p(\mathbf{x}, \Omega) = \frac{\det(\Omega)^{1/2}}{2^{N/2}} \mathrm{e}^{-(\mathbf{x}-\boldsymbol{\mu})^\top \Omega (\mathbf{x}-\boldsymbol{\mu})/2}$$

$$\Longrightarrow \quad \log(p(\mathbf{x}, \Omega)) = \frac{1}{2} \log\det(\Omega) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Omega (\mathbf{x}-\boldsymbol{\mu}) + c$$

Smith, Stephen M., et al. "Network modelling methods for FMRI." *Neuroimage* ,vol. 54, no. 2, pp. 875-891, 2011.

Regularized inverse covariance

$$\min_{\Omega \geq \mathbf{0}} - \log\det(\Omega) + \mathrm{tr}(\Omega\Sigma) + \lambda\|\Omega\|_1$$

$$\|\Omega\|_1 = \sum_{i=1}^{N}\sum_{j=1}^{N}|\omega_{ij}|$$

Through a seq. of equiv. problems $\Downarrow$

$$\max_{\Omega \geq \mathbf{0}} \log\det(\Omega)$$

s.t. $\|\Omega - \Sigma\|_\infty \leq \lambda$

Diagonal entries

$$\widehat{w}_{ii} = \max_{|w_{ii} - s_{ii}| \leq \lambda} w_{ii} = s_{ii} + \lambda$$

Off-diagonal entries

cyclic minimization w.r.t. $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_N$

without touching diagonal entries

$$\boldsymbol{\omega}_i^{(k+1)} = \arg\max_{\boldsymbol{\omega}_i} \log\det(\boldsymbol{\Omega})$$

$$\text{s.t. } \|\boldsymbol{\omega}_i - \mathbf{s}_i\|_\infty \leq \lambda$$

equivalent to LASSO

O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data," *J. Machine Learning Research*, vol. 9, pp. 485-516, June 2008.

Granger causality and lag-based methods

- **Main idea:** Cause precedes effect

A causes B $\Longrightarrow$ knowing past of A helps the prediction of B (than only knowing past of B)

- **Implementation**: Multivariate autoregressive modelling

- Different variations/toolboxes available
  - "Casual connectivity analysis" toolbox
    - measures <u>conditional</u> granger causality, needs specification of model order
  - "Bayesian information criterion"
    - Estimates model order up to a specified maximum
  - "BioSig" toolbox

S. M. Smith, et al. "Network modelling methods for FMRI." *Neuroimage* ,vol. 54, no. 2, pp. 875-891, 2011.

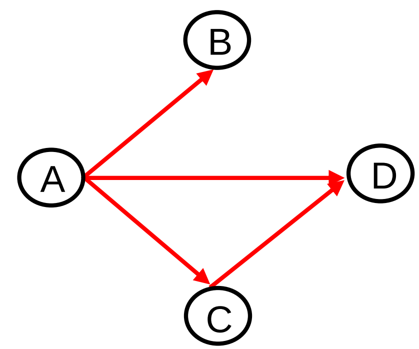
Structural Equation Models (SEM)

- Structural Equations

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix}$$

$$\begin{array}{c|cccc|} & \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} \\ \hline \mathrm{A} & 0 & 0 & 0 & 0 \\ \mathrm{B} & 1 & 0 & 0 & 0 \\ \mathrm{C} & 1 & 0 & 0 & 0 \\ \mathrm{D} & 1 & 0 & 1 & 0 \\ \hline \end{array}$$

- Comparison of two models with/without a connection
- Prevention of model order growth (Bayesian, etc.)
- Search methods (exhaustive, greedy)
- A prior anatomical models used as starting graphs



- Hard to model latent variables in the model in the presence of fMRI noise

Dynamic Causal Models (DCM)

Objective

Discover the causal architecture of coupled or distributed dynamic systems

Main Idea

- Estimate interactions among brain regions in experimental changes
- Map the neural activity (hidden states) to measured response

$$\mathbf{u} \rightarrow \mathbf{x} \rightarrow \mathbf{y}$$

Method

- Model the system as differential equations
- Perform Bayesian model comparison for model selection
- Characterize model in terms of its parameters (parameter estimation)

Andre C. Marreiros et al. (2010), Scholarpedia, 5(7):9568.

Dynamic Causal Models (DCM)

A. Modeling (deterministically) the changes of a neural state-vector $\mathbf{x}$ in time:

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \theta) = \mathbf{A}\mathbf{x} + \sum_{j=1}^{m} u_j \mathbf{B}^{(j)}\mathbf{x} + \mathbf{C}\mathbf{u}$$

$$\mathbf{A} = \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{u}=0}$$

Fixed (average) coupling

$$\mathbf{B} = \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{u}}$$

Effective change in the coupling induced by the j-th input

$$\mathbf{C} = \left.\frac{\partial f}{\partial \mathbf{u}}\right|_{\mathbf{x}=0}$$

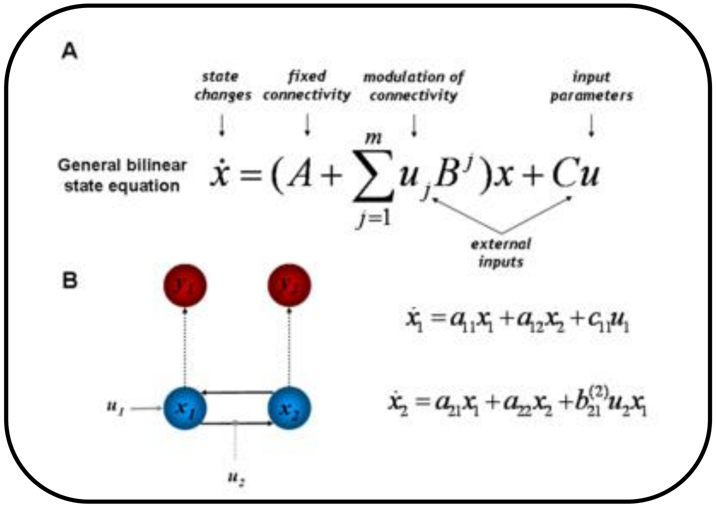
Influence of input that cause perturbation of hidden state

- Simplest low-order approximation accounting for endogenous-exogenous causes

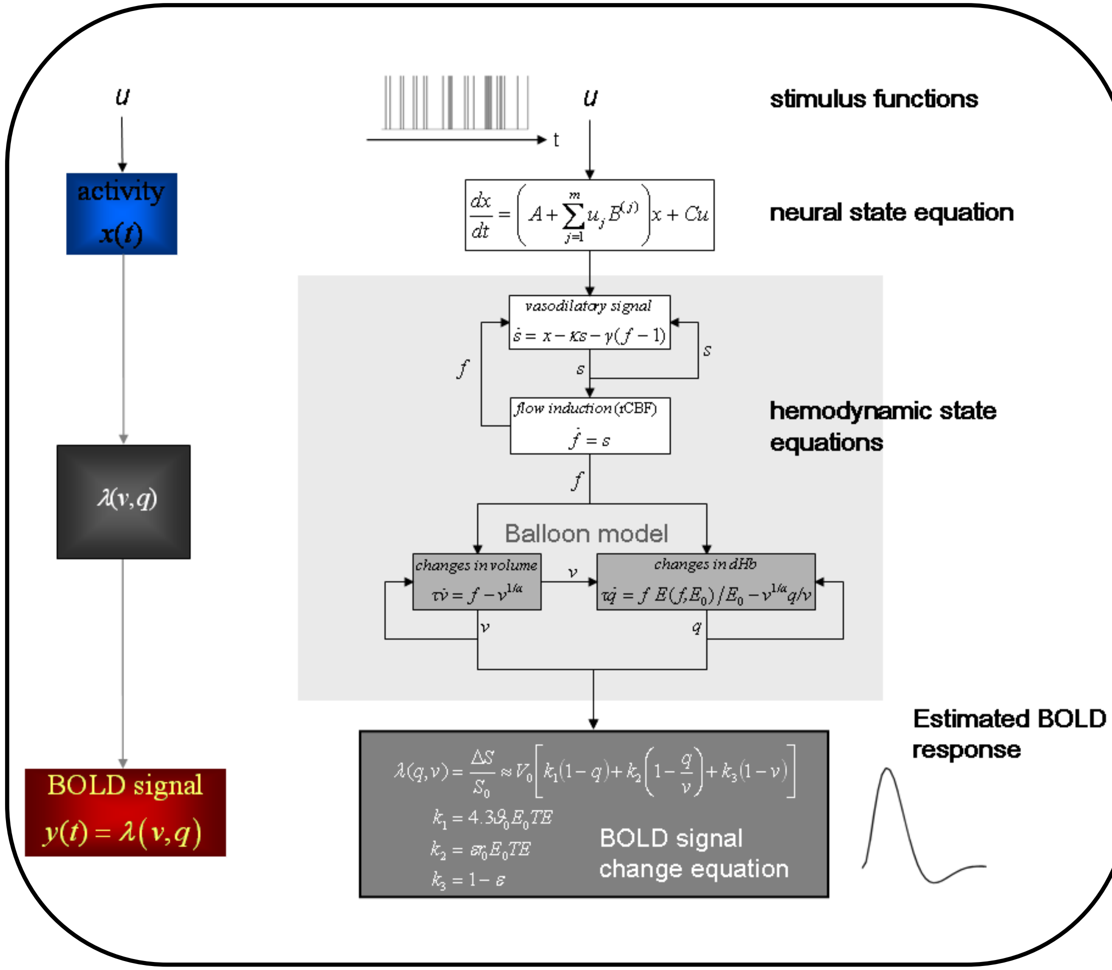
B. Transformation of neural-activity into BOLD responses $\Rightarrow$ The Balloon Model (Buxon *et al.*, 1998)

neural-activity $\rightarrow$ Blood flow and volume $\rightarrow$ Deoxyhemoglobin content

DCM + The Balloon model



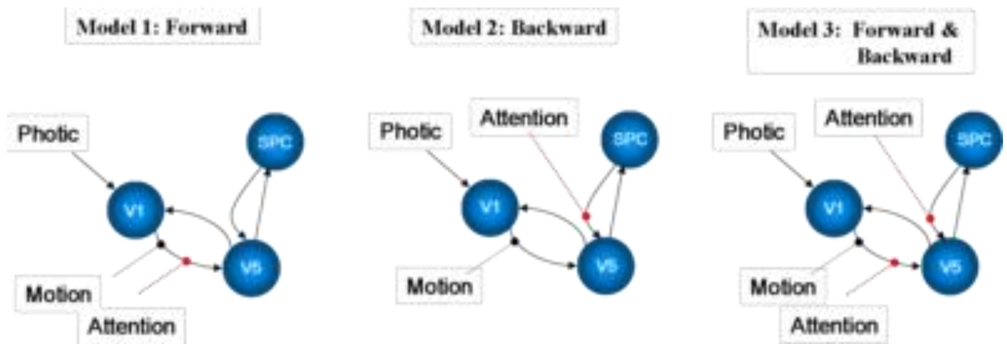
(Stephan *et al.*, 2007a)



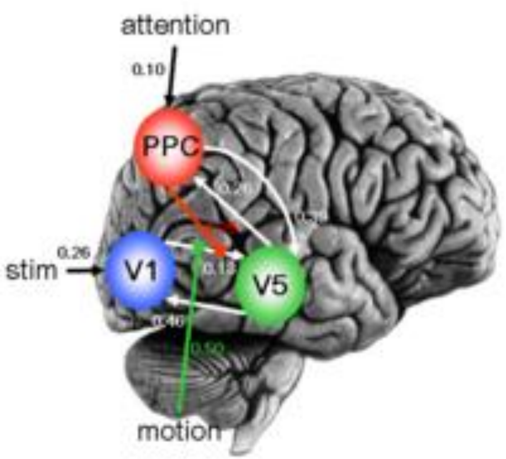
(Friston *et al.*, 2003)

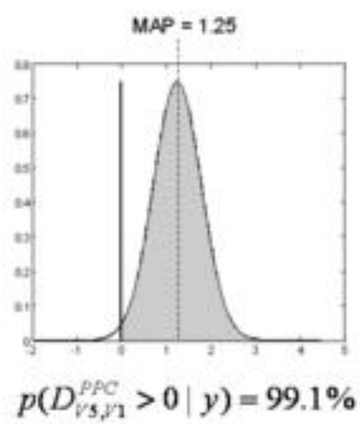
DCM- fMRI Application

DCM applied to data from a study on attention to visual motion. In all models, photic stimulation enters V1 and motion modulates the connection from V1 to V5. They differ in how attention (red) modulates the connectivity to V5; with model 1 assuming modulation of the forward connection (V1 to V5), model 2 assuming modulation of the backward connection (SPC to V5) and model 3 assuming both.



(Penny *et al.*, 2004)





(Stephan *et al.*, 2008).

Nonlinear DCM for fMRI applied to the attention to motion paradigm. Left panel: Numbers alongside the connections indicate the *maximum a posteriori* (MAP) parameter estimates. Right panel: Posterior density of the estimate for the nonlinear modulation parameter for the V1→V5 connection. Given the mean and variance of this posterior density, we can be 99.1% confident that the true parameter value is larger than zero or, in other words, that there is an increase in gain of V5 responses to V1 inputs that is mediated by parietal activity.

Other methods

- Mutual Information (MI)
- Coherence: Dependence in the frequency domain
- Bayes Nets
- LiNAG

Recap

Thank You!

- Graphical modeling of brain networks
  - Graph vertex assignment
    - ICA
    - PCA
    - Clustering
  - Vertex time-series assignments
  - Graph edge assignment
    - Correlation and partial correlation
    - Inverse covariance matrix
    - Granger causality
    - SEM
    - CDM
- Structural vs. functional vs. effective connectivity in brain