Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz Association

Connectivity networks in neuroscience - construction and analysis

Jörg Polzehl

Content

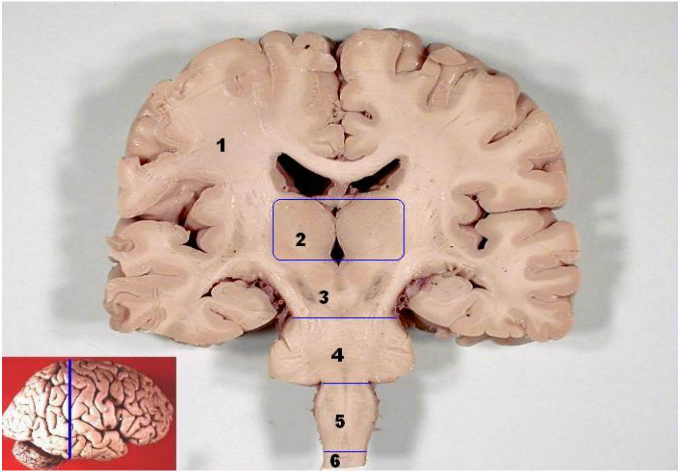


Figure: John A Beal, PhD Dep't. of Cellular Biology & Anatomy, Louisiana State University Health Sciences Center Shreveport

(Wikimedia)

Brain organisation:

- Gray matter / Cortex
- White matter
- CSF (Cerebrospinal fluid)

Cortex divided in

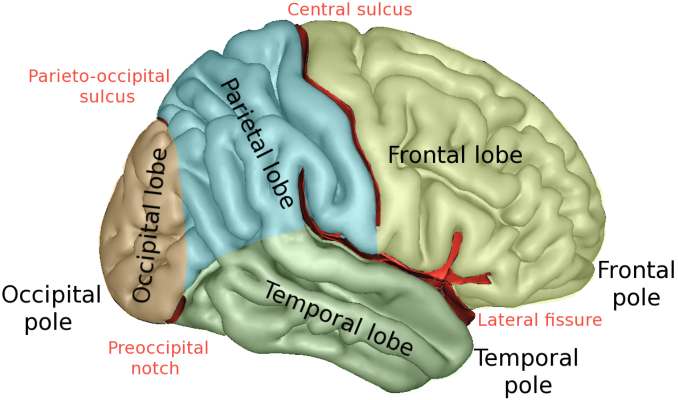
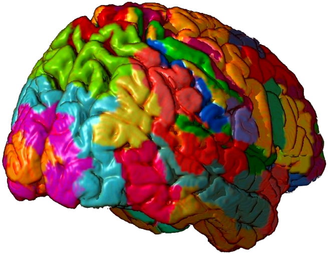


Figure: Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436.

Lobes, Gyri (ridge on the cerebral cortex), Sulci (depression or groove in the cerebral cortex)

White Matter: Fiber bundles connecting cortical areas



Brodman Areas:

- ≈ 50 cortex regions defined based on cytoarchitectural organization of neurons
- regions have been correlated to cortical functions

Brodmann K (1909). ”Vergleichende Lokalisationslehre der Grosshirnrinde”. Leipzig: Johann Ambrosius Barth

Figure: Mark Dow. Research Assistant Brain Development Lab, University of Oregon.

http://lcni.uoregon.edu/ dow/Space_software/renderings.html

Talaraich atlas: single subject

Talairach et al. Co-planar stereotaxic atlas of the human brain. Thieme, New York. (1988)

Problem: Subject variability

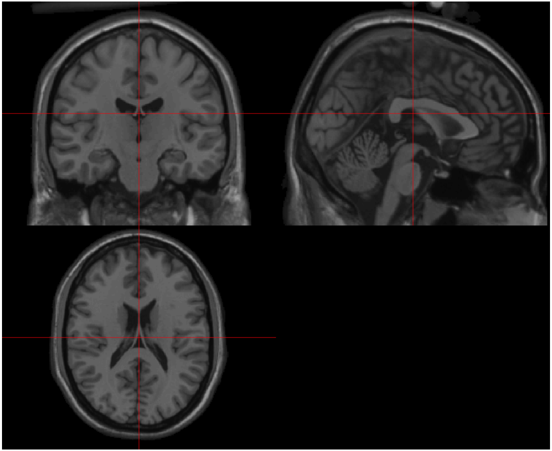
Talaraich atlas: single subject

Talairach et al. Co-planar stereotaxic atlas of the human brain. Thieme, New York. (1988)

Problem: Subject variability

MNI (Montreal Neurological Institute) template (ICBM152): Average of 152 MRI scans matched by affine transform (9 parameters)

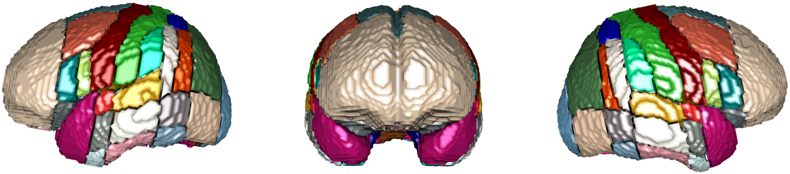
Maintained by the International Consortium for Brain Mapping



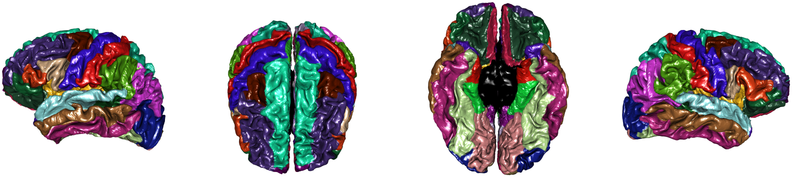
Orthographic view of MNI template

Probabilistic brain atlases

Havard-Oxford atlas (FSL) : 48 cortical + 21 subcortical regions, 37 subjects



Probabilistic atlas used in FSL (Desikan et al., (2006). NeuroImage, 31(3):968-80.)



Partially addresses subject variability

- Term covers a number of minimally invasive techniques to study the brain
- used to characterize structure, function and diagnostic of diseases
- contribute to understanding interactions between mind (decisions, emotions), brain and body

Two categories

Structural neuroimaging

Functional neuroimaging

- Term covers a number of minimally invasive techniques to study the brain
- used to characterize structure, function and diagnostic of diseases
- contribute to understanding interactions between mind (decisions, emotions), brain and body

Two categories with modalities

Structural neuroimaging

- Computed tomography (CT)
- Positron emission tomography (PET)
- Magnetic resonance imaging (MRI)
- Diffusion weighted magnetic resonance imaging (dMRI/DWI)

Functional neuroimaging

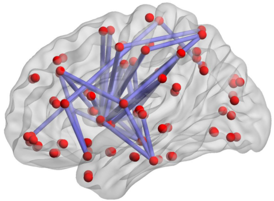
- Electroencephalography (EEG)
- Magnetoencephalography (MEG)
- Positron emission tomography (PET)
- functional magnetic resonance imaging (fMRI)

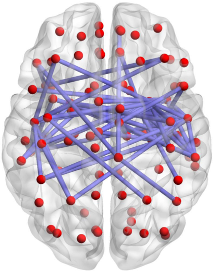
Describes the interaction of cortical brain regions

- Functional connectivity: characterises the simultaneous function of different brain regions
- Structural (anatomic) connectivity: describe the anatomical connection of functional brain regions (nodes) by white matter fiber tracks
- Effective connectivity: describe the causal interaction of functional brain regions by directed graphs

All require definition of nodes (functional regions) by some methods:

- use of anatomic information (cortical thickness, myelination)
- functional regions identified by fMRI experiments
- default networks identified by resting state fMRI experiments





Source: Wikimwdia

Magnetic resonance imaging (MRI)

Figure: Kasuga Huang (Wikimedia)

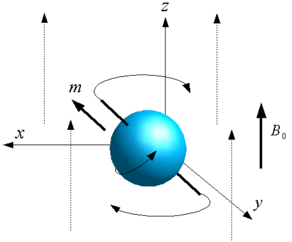


Figure: Franz Wilhelmstötter (Wikimedia)

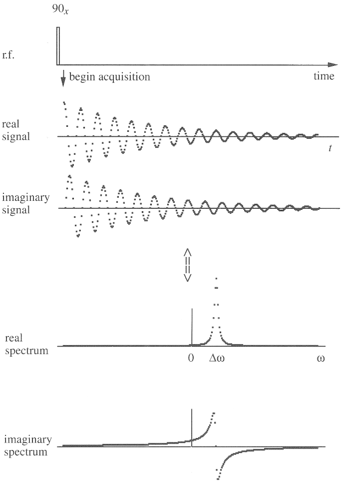


Fig. 2.7 Free Induction Decay (F I D) following a single 90° r.f. pulse. The real and imaginary parts of the signal correspond to the in-phase and quadrature receiver outputs. The signal is depicted with receiver phase $\phi = 0$ and, on complex Fourier transformation, gives real absorption and imaginary dispersion spectra at the offset frequency, $\Delta\omega = \omega_0 - \omega$.

From O. Friman "Adaptive Analysis of Functional MRI Data", PhD Thesis, 2003

Figure: Kasuga Huang (Wikimedia)

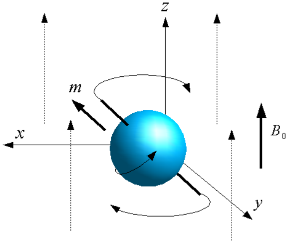


Figure: Franz Wilhelmstötter (Wikimedia)

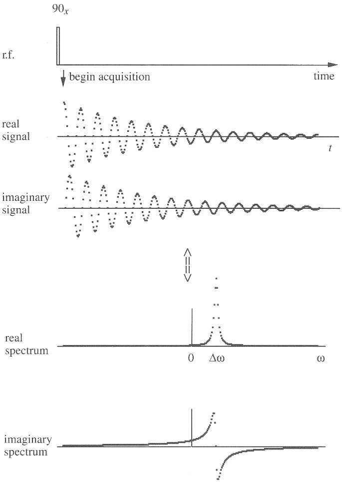
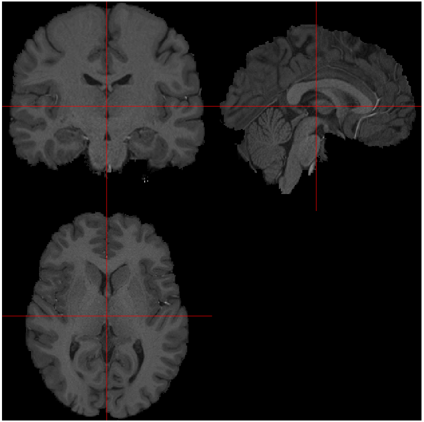
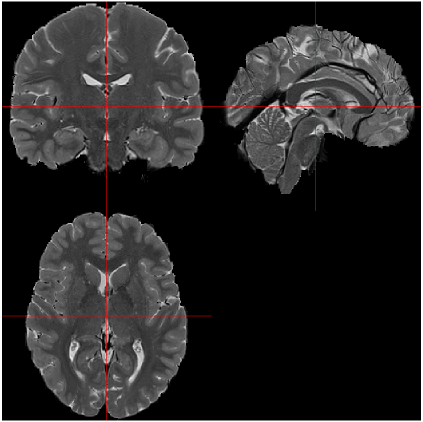


Fig. 2.7 Free Induction Decay (F I D) following a single 90° r.f. pulse. The real and imaginary parts of the signal correspond to the in-phase and quadrature receiver outputs. The signal is depicted with receiver phase $\phi = 0$ and, on complex Fourier transformation, gives real absorption and imaginary dispersion spectra at the offset frequency, $\Delta\omega = \omega_0 - \omega$.

From O. Friman “Adaptive Analysis of Functional MRI Data”, PhD Thesis, 2003



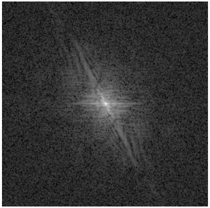
T1-weighted image (orthographic view), longitudinal relaxation



T2-weighted image (orthographic view), transverse relaxation

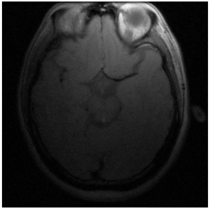
[0]Data provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

k-space



FFT

image-space



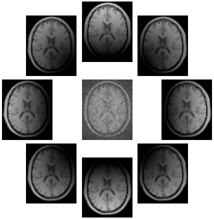
GRAPPA

SENSE

SMASH

$\dots$

combined magnitude



Complex Gaussian
$F_c(k_x, k_y)$
multiple receiver coils c

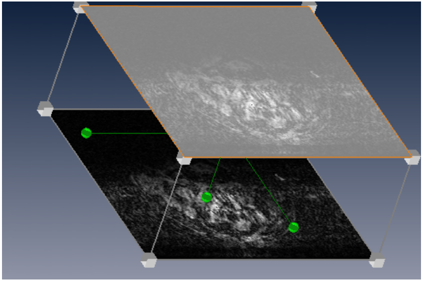
Complex Gaussian $f_c(x, y)$
Rician magnitude image
$Mod(f_c(x, y))$

$\frac{S}{\sigma} \sim \chi_{2L^*, \zeta/\sigma}$
Non-central χ
distribution

- 24-32 receiver coils
- Acquisition protocol and reconstruction method cause (non-local) spatial correlation
- Signal distribution depends on the reconstruction method (SENSE, GRAPPA, SoS, ...)

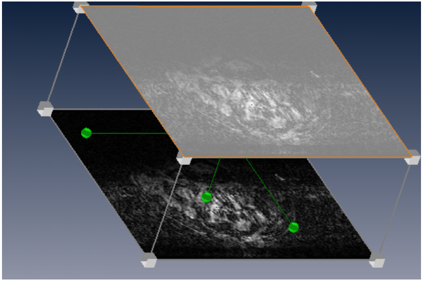
Structural MR images:

- high spatial resolution
- offer contrast between tissue types (cortex <–> white matter)
- no temporal information



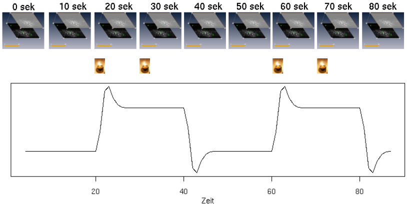
Structural MR images:

- high spatial resolution
- offer contrast between tissue types (cortex <–> white matter)
- no temporal information



Functional MR images

- lower spatial resolution
- lower image contrast
- temporal resolution
- signal changing with experimental tasks



Uses the Blood Oxygenation Level Dependent (BOLD) contrast

- Active neurons need oxygen!
- Change of magnetic properties due to oxygenation.
- Measure the ratio of oxygenated to deoxygenated hemoglobin
- Local signal changes over time due to brain function

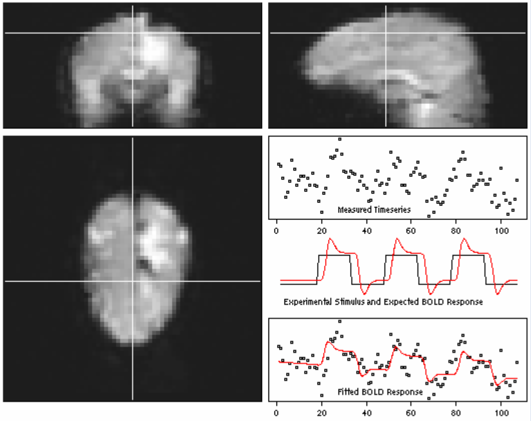
Experiment:

- Measurement of fast time series of the brain under stimulus

Indirect measurement:

- Measures oxygen consumption of active neurons
- Signal changes are delayed in time
- Convolution with hemodynamic response function
- Limited spatial resolution by vascular architecture

fMRI data = 3D + T



About fMRI data

- Time series of 3D data
- Spatial resolution: 1-4 mm
- Temporal resolution: 1-3 sec
- Search for locations, were a BOLD signal can be found!
- Problem: noise
- Problem: multiple test problem

- Realignment/Registration:
  - Corrections for head movement
  - Rigid or affine transformation
- Slice time correction:
  - Adjust for slice recording at different times
- Normalization:
  - Mapping to a standard space (Talaraich, MNI)
  - Comparability between subjects in group studies
- Cortex segmentation based on corresponding anatomic images
- Spatial smoothing

Hemodynamic response function:

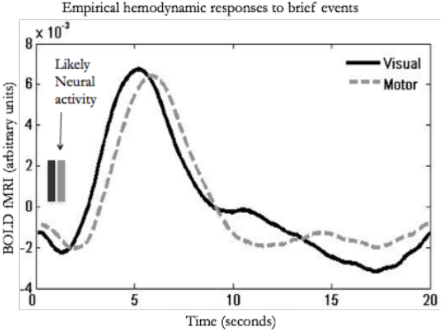


Figure: M. Lindquist, J. Hopkins Univ., Talk at SAMSI 2015
Study: Lindquist et al., Journal of Magnetic Resonance 2008

Parametric model:

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} \exp\left(-\frac{t-d_1}{b_1}\right) - c\left(\frac{t}{d_2}\right)^{a_2} \exp\left(-\frac{t-d_2}{b_2}\right)$$

Time delay modeled by including the derivative of h

Spatially varying form and latency

Expected BOLD response: Convolution between stimulus and hemodynamic response

Linear model:

$$Y_i = X\beta_i + \varepsilon_i$$

- Data $Y_i = (Y_{it})$
- Design $X_i = (x_{itk})$, i - voxel, t-time, $k = 1, K$ - components
- Error $\varepsilon_i = (\varepsilon_{it})$, $\mathbf{E}\,\varepsilon_{it} = 0$, $\mathbf{E}\,\varepsilon_{it}^2 = \sigma_t^2$, $\mathbf{Cov}(\varepsilon_{it}, \varepsilon_{i(t-j)}) = \delta_{ij}$, usually $AR(1)$ or $AR(2)$

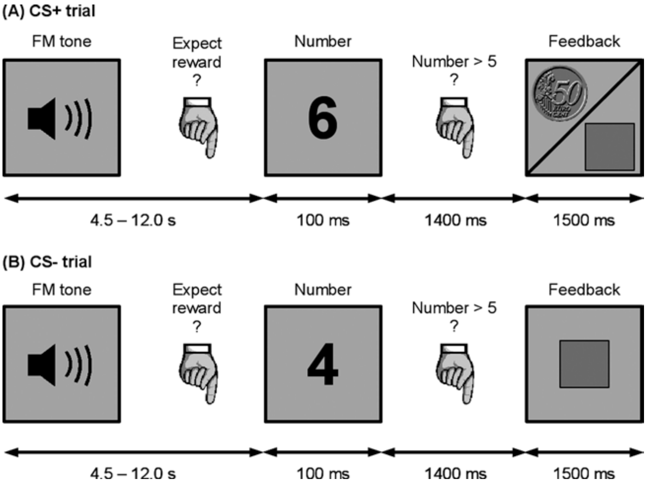
Components include:

- Expected bold responses to stimuli
- drift components for magnetic field inhomogeneity (polynomial)
- confounding (physiological) effects (respiration, cardiac cycle, ...)
- parameters from motion correction

Prewhitening: Transform model such that errors are approx. uncorrelated

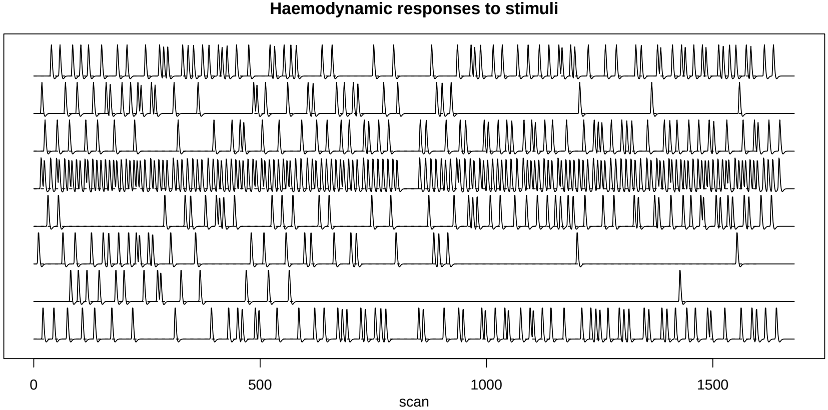
$$\tilde{Y}_i = \tilde{X}_i\beta_i + \tilde{\varepsilon}_i, \quad \tilde{Y}_i = A_iY_i, \quad \tilde{X}_i = A_iX, \quad \tilde{\varepsilon}_i = A_i\varepsilon_i,$$

Learning paradigm:



(Figure: Puschmann (2013))

Stimulus components in design matrix:



Interest in contrast

$$\gamma = c^T \beta$$

and testing

$$H : \gamma_i = 0 \quad \text{against} \quad A : \gamma_i <> 0$$

to determine active brain regions associated with the contrast

- Estimate $AR(k)$ parameters from residuals in linear model $Y_i = X\beta_i + \varepsilon_i$
- Spatially smooth $AR(k)$ parameters
- Prewhitening using $\hat{A}_i$ obtained from smoothed $AR(k)$ parameters
- Estimation of β_i:

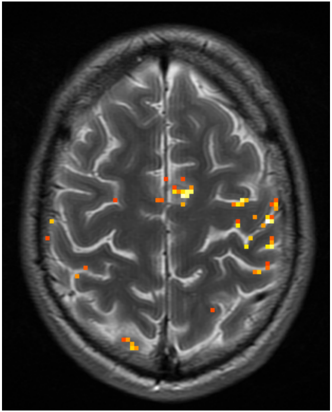
$$\hat{\beta}_i = \left(\tilde{X}_i^T \tilde{X}_i\right)^{-1} \tilde{X}_i \tilde{Y}_i$$

- Estimate covariance $\hat{\Sigma}_i$ of $\hat{\beta}_i$ from prewhitened model
- Define test statistics (t-distributed)

$$S_i = \frac{c^T \hat{\beta}_i}{(c^T \Sigma_i c)^{1/2}}$$

Multiple test problem

- Simultaneous tests in $N = 10000$ (Cortex) - 100000 (Brain) voxel
- using t-thresholds at significance level α gives $\approx \alpha N$ false positives.
- Adjustment for multiple testing by Bonferroni leads to high thresholds
- Multiplicity adjustment leads to low sensitivity
- Alternative: False Discovery Rate (Benjamini & Hochberg 1995)
  Control of proportion of false positives within detected signals
- ignores spatial extend of regions of interest



voxelwise decision using thresholds adjusted for multiple testing

- Regions of activation have a spatial extend
- Smoothing the observed images with a (Gaussian) kernel with bandwidth h

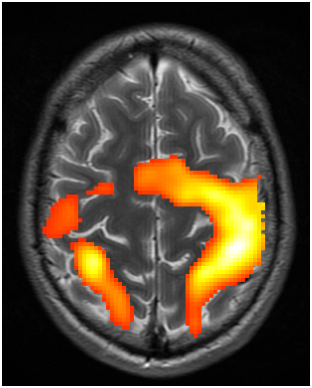
$$\bar{Y}_{it} = \sum_j K\left(\frac{||i-j||}{h}\right) Y_{jt}$$

decreases variance and increases Signal-to-Noise ratio (SNR)

- reduce the number of independent decisions.
- thresholds can be obtained by Random Field Theory (Adler 1987, 2000, Worsley 1994ff)

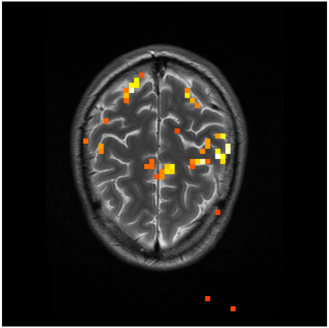
$$P(\max_i \bar{S}_i > \tau) \approx \sum_{d=0}^{3} R_d(V,h)\rho_d(\tau)$$

$R_d(V,h)$ - d-dimensional resel count
ρ_d- d-dimensional Euler characteristic density.



decision using nonadaptive smoothing and thresholds given by Random Field Theory

Results of a voxelwise analysis,



Motor (finger tapping task)

- Interpretation of test results ?
  - Poor signal to noise
  - Activation or other sources of variation (motion artifact, physiological noise, other processes ?
  - Color coded p-values
  - low sensitivity <–> reduced spatial resolution
  - Search for activated regions instead of activated voxel
- Reproducibility of results ??
  - large variability over repeated experiments (same subject)
  - Representativity for populations ?
  - Between subject variability
  - Group studies needed

fMRI experiments without external stimulus (resting state)

- looks for intrinsic brain activity
- first experiments by Biswal et al. (1995) observe patterns of spatial coherence between sensorimotor regions
- e.g. Zang & Raichle (2010) identify 7 mayor networks of regions that show spatially coherent activity
- larger studies identify up to 17 networks

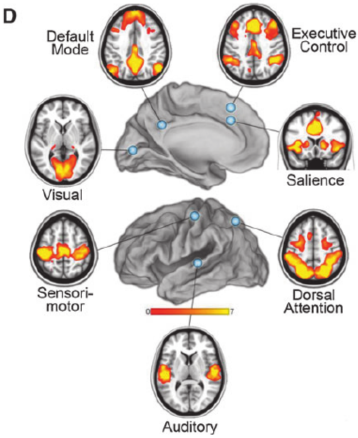


Figure: Raichle, Brain Connectivity, 2011, Fig. 1D

Independent component analysis (ICA)

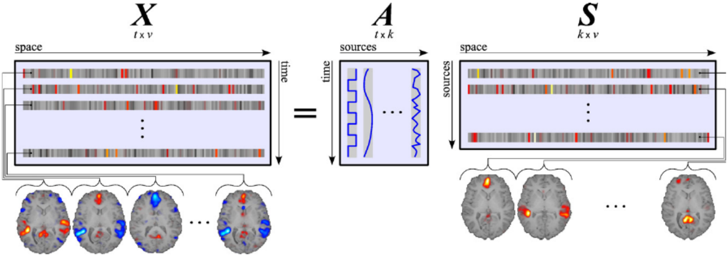
- Observed signals $x_1(t), x_2(t), \ldots, x_p(t)$
- Assume these signals to be a linear combination of unknown souces $s_1(t), s_2(t), \ldots, s_q(t)$
- Model:

$$x_i(t) = \sum_{k=1}^{K} a_{ik} s_k(t) + \varepsilon_i(t) \quad i = 1, \ldots, p \tag{1}$$

$$X = AS + E \tag{2}$$

- Goal: Estimate the mixing matrix $A = (a_{ij})$ and the unknown source signals $s_j(t)$
- Source separation or cocktail party problem

- Data: $n_1 \times n_2 \times n_3 \times T$ values. Reorder as data matrix $n \times T$
- Reduction of data matrix by Prewhitening and PCA, specification of number of sources K
- Search for spatial pattern in S (Spatial ICA)



Ylipaavalniemi and Vigário, Neuroimage 2008

- Decompose in temporal (A) and spatial S signals
- Solved by e.g. fastICA (Hyvärinen & Oja (2000))
- Some components k may model artifacts (interpretability !)

- C.F. Beckmann and S.M. Smith, Neuroimage 2005
- Generalization of ICA for group studies
- K subjects
- Model:

$$X_{IK\times J} = (C|\otimes|A)S + E_{IK\times J} \tag{3}$$

$$(C|\otimes|A) = ((A\,\mathrm{diag}(c_1))^\top, \ldots, (A\,\mathrm{diag}(c_K))^\top)^\top \tag{4}$$

- Structure of mixing matrix $(C|\otimes|A)$ reflects the individual effects
- Common spatial structure in S

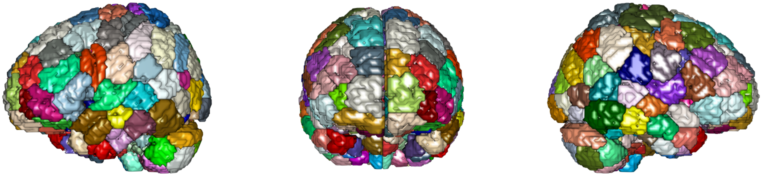
- 1200 Subjects
- anatomical scans $0.7\mathrm{mm}$ isotropic (T1/T2)
- task based fMRI, 7 tasks $2\mathrm{mm}$ isotropic
  (Working memomory, Gambling, Motor, Language, Social cognition, Relational processing, Emotion Processing)
- resting state fMRI $4 \times 15\mathrm{min}$ $2\mathrm{mm}$ isotropic
- diffusion weighted imaging $1.25\mathrm{mm}$ isotropic, 3×90 gradients

Information from these experiments is combined to obtain individual brain parcellations (node definitions) for all subjects

Literature:

- Special issue Neuroimage 2013
- Shen et al. Neuroimage 2013
- Finn et al., Nature neuroscience 2015

Brain parcellation, 268 functional regions, Shen 2013



- Finn 2015 defines general procedure for corresponding subject specific region definition
- regions should be used for node definition in group studies

- Selection of characteristic time series within regions
- leads to matrix $Y = (y_{kt})_{k=1,K}^{t=1,T}$
- define network by empirical covariance matrix

$$\hat{\Sigma} = \left(\sum_t (y_{it} - y_{i.})(y_{jt} - y_{j.}) \right)_{i,j=1,K}$$

- or regularized / thresholded estimate

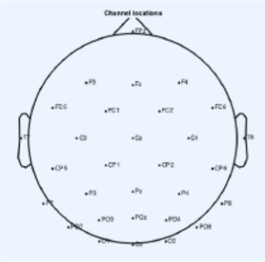
Task based fMRI depending on the goal:

- Modeling and removal of expected hemodynamic response
- Selection of characteristic (residual) time series within regions

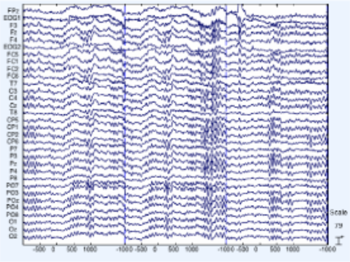
or

- Selection of nodes using functional regions associated with the tasks

EEG channel locations



EEG signals at each location



Source: M. Lindquist, J. Hopkins Univ., Talk at SAMSI 2015

- high temporal resolution
- low spatial resolution
- Indirect measurement
- Source reconstruction problem
- Networks: coherence between spectra of recorded or reconstructed signals

Lit: Ombao and Van Bellegem (2008). Coherence Analysis: A Linear Filtering Point Of View. IEEE Transactions on Signal Processing, 56(6), 2259-2266.

Assumptions:

$$Y_t \sim N_p(0, \Sigma), \qquad \Sigma = (\sigma_{ij})_{i,j=1}^p$$

Correlation between signals in nodes (regions) describes joint activity

$$R = (\rho_{ij})_{i,j=1}^p, \qquad \rho_{ij} = \frac{\sigma_{ij}}{(\sigma_{ii}\sigma_{jj})^{1/2}}$$

Assumptions:

$$Y_t \sim N_p(0, \Sigma), \qquad \Sigma = (\sigma_{ij})_{i,j=1}^p$$

Correlation between signals in nodes (regions) describes joint activity

$$R = (\rho_{ij})_{i,j=1}^p, \qquad \rho_{ij} = \frac{\sigma_{ij}}{(\sigma_{ii}\sigma_{jj})^{1/2}}$$

Partial correlations refer to joint activity not explained by intermediate effects

$$P = (\rho_{ij.k})_{i,j=1}^p, \qquad \rho_{ij.k} = \frac{\sigma_{ij} - \sigma_{ik}^T \Sigma_k^{-1} \sigma_{jk}}{((\sigma_{ii} - \sigma_{ik}^T \Sigma_k^{-1} \sigma_{ik})(\sigma_{jj} - \sigma_{jk}^T \Sigma_k^{-1} \sigma_{jk}))^{1/2}}$$

with $k = (1 \ldots n)/(ij)$

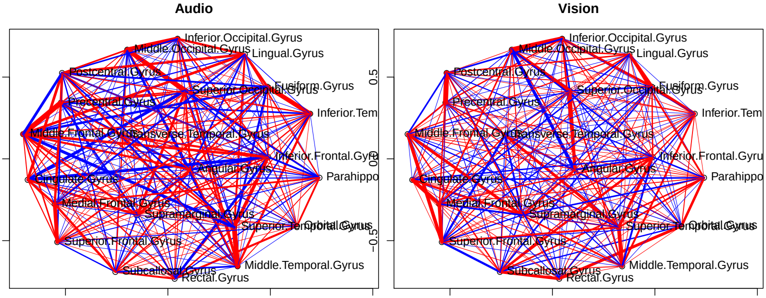
Precision matrices: $\Omega = \Sigma^{-1} = (\omega_{ij})_{i,j=1}^p$,

Connection to partial correlations: $\rho_{ij.k} = -\frac{\omega_{ij}}{(\omega_{ii}\omega_{jj})^{1/2}}$

- Pourahmadi, M.: Modeling covariance matrices: The GLM and regularization perspectives. Statist. Sci., 2011, 26,, 369-87.

Negative normal log-likelihood: $Y_t \sim N_p(0, \Sigma)$, $S = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \bar{Y})(Y_t - \bar{Y})^T$

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \log|\Omega| - tr(S\Omega), \qquad \hat{\Omega} = S^{-1}$$



- functional connectivity networks are hypothesized to be sparse
- $p = 22$, $n = 178$ –> high variability of estimated correlations

Regularization:

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \log|\Omega| - tr(S\Omega) + \mathscr{P}_\lambda(\Omega)$$

Graphical LASSO:

$$\mathscr{P}_\lambda(\Omega) = \lambda \sum_{ij}^{p} |\omega_{ij}|$$

Literature:

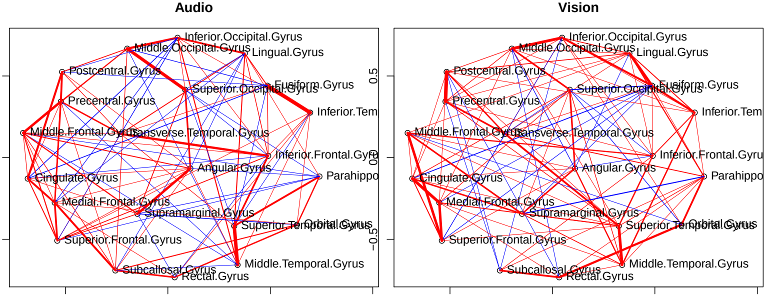
- Meinshausen, N. & Bühlmann, P.: High-dimensional graphs and variable selection with the Lasso, Ann. Stat., 2006
- Friedman, J.; Hastie, T. & Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, Biostatistics, 2008
- Levina, E.; Rothman, A. J. & Zhu, J.: Sparse estimation of large covariance matrices via a nested lasso penalty, Ann. Appl. Stat., 2008
- Rothman, A. J.; Levina, E. & Zhu, J.: Generalized Thresholding of Large Covariance Matrices, JASA, 2009
- Rothman, A. J., L. E. & Zhu, J.: Sparse multivariate regression with covariance estimation, JCGS, 2010
- Bien, J. & Tibshirani, R.: Sparse Estimation of a Covariance Matrix, Biometrika, 2011
- Rothman, A. J.: Positive definite estimators of large covariance matrices Biometrika, 2012
- Mazumder, R. & Hastie, T.: The Graphical Lasso: New Insights and Alternatives, Electr. J. Stat., 2012
- Mazumder, R. & Hastie, T.: Exact covariance thresholding into connected components for large-scale Graphical Lasso, JMLR, 2012

Regularization:

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \log|\Omega| - tr(S\Omega) + \mathscr{P}_\lambda(\Omega)$$

Graphical LASSO:

$$\mathscr{P}_\lambda(\Omega) = \lambda \sum_{ij}^{p} |\omega_{ij}|$$



Solution for $\lambda = .1$ function dpglasso from R-package dpglasso.

Problem: Produces a biased estimate !

Regularization:

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \log|\Omega| - tr(S\Omega) + \sum_{ij}^{p} p_\lambda(\omega_{ij})$$

adaptive LASSO (Hui Zou):

$$p_\lambda(\omega_{ij}) = \lambda \frac{1}{\tilde{\omega}_{ij}^{\gamma}} |\omega_{ij}|$$

SCAD (Smoothly Clipped Absolute Deviation) (Fan & Li (2001)):

$$p_\lambda(\omega_{ij}) = (\lambda I_{|\tilde{\omega}_{ij}| \leq \lambda} + \frac{(a\lambda - |\tilde{\omega}_{ij}|)_+}{(a-1)} I_{|\tilde{\omega}_{ij}| > \lambda}) |\omega_{ij}|$$

Suggested parameters: $\gamma = .5$, $a = 3.7$. $\tilde{\omega}$ are assumed to be consistent estimates.

Computations:

- Non-convex optimization problems
- can be approximated by iteration of graphical LASSO (with matrix penalty parameter)

- Zou, H.: The Adaptive Lasso And Its Oracle Properties, JASA, 2006, 101, 1418-1429
- Lam, C. & Fan, J.: Sparsistency and Rates of Convergence in Large Covariance Matrices Estimation, Ann. Stat, 2009
- Fan, J.; Feng, Y. & Wu, Y.: Network exploration via the adaptive LASSO and SCAD penalties, Ann. Appl. Stat, 2009
- Cai, T. T.; Liu, W. & Zhou, H.: Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation, Ann. Stat., 2014.
- Cai, T.; Liu, W. & Luo, X.: A Constrained l1 Minimization Approach to Sparse Precision Matrix Estimation, JASA, 2011

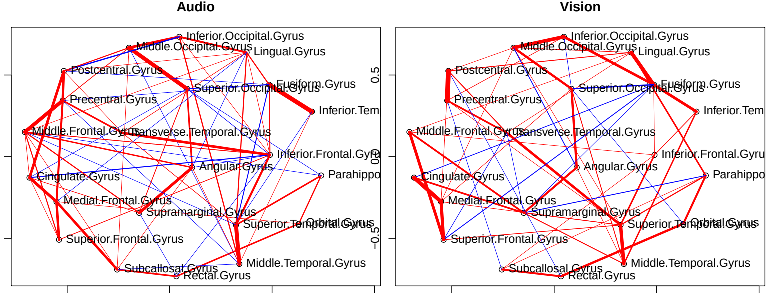
Adaptive penalties (adaptive LASSO)

Regularization:

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \log|\Omega| - tr(S\Omega) + \sum_{ij}^{p} p_\lambda(\omega_{ij})$$

adaptive LASSO (Hui Zou):

$$p_\lambda(\omega_{ij}) = \lambda \frac{1}{\tilde{\omega}_{ij}^{\gamma}} |\omega_{ij}|$$



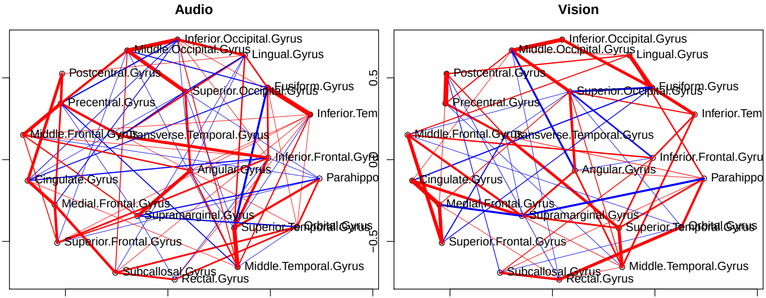
Parameters: $\lambda = .1$, $\gamma = .5$.

Regularization:

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \log|\Omega| - tr(S\Omega) + \sum_{ij}^{p} p_\lambda(\omega_{ij})$$

SCAD (Fan & Li (2001)):

$$p_\lambda(\omega_{ij}) = (\lambda I_{|\tilde{\omega}_{ij}|\leq\lambda} + \frac{(a\lambda - |\tilde{\omega}_{ij}|)_+}{(a-1)} I_{|\tilde{\omega}_{ij}|>\lambda})|\omega_{ij}|$$



Parameters: $\lambda = .1$, $a = 3.7$.

Proposals (based on model selection criteria) with $\Lambda = (\lambda_{ij})$

- K-fold - Cross-validation

$$KCV(\Lambda) = \sum_{k=1}^{K} n_k (\log|\hat{\Omega}^{(-k)}(\Lambda)| - tr(S^{(k)}\hat{\Omega}^{(-k)}(\Lambda))$$

- Generalized Cross validation (Dong & Wahba 1996, Lian 2011)

$$\begin{aligned} GACV(\Lambda) =& n(\log|\hat{\Omega}(\Lambda)| - tr(S\hat{\Omega}(\Lambda)) + \\ &+ \sum_{i=1}^{n} vec(\hat{\Omega}(\Lambda)^{-1} - y_i y_i^T)^T vec(\hat{\Omega}(\Lambda)(S^{(-i)} - S)\hat{\Omega}(\Lambda)) \end{aligned}$$

- Bayes Information Criterion (BIC) (consistent !)

$$BIC(\Lambda) = -\log|\hat{\Omega}(\Lambda)| + tr(S\hat{\Omega}(\Lambda)) + k\frac{log(n)}{n}$$

Suggestion: select maximum λ such that BIC slightly exceeds its minimal value.

- Lian, H.: Shrinkage tuning parameter selection in precision matrices estimation, J. Stat. Plan. Inf., 2011
- Chatterjee, A. & Lahiri, S. N.: Bootstrapping Lasso Estimators, JASA, 2011

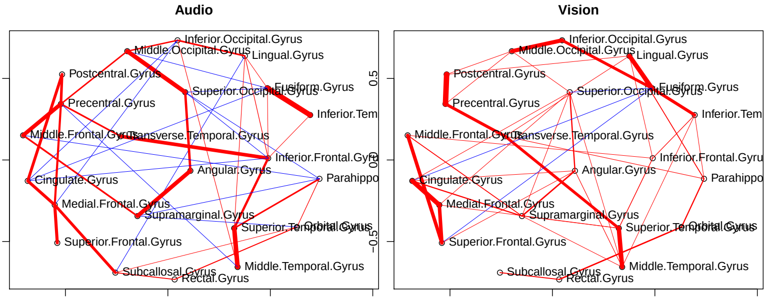
SCAD with λ chosen by BIC

SCAD (Fan & Li (2001)):

$$p_{\lambda}(\omega_{ij}) = (\lambda I_{|\tilde{\omega}_{ij}| \leq \lambda} + \frac{(a\lambda - |\tilde{\omega}_{ij}|)_{+}}{(a-1)} I_{|\tilde{\omega}_{ij}| > \lambda})|\omega_{ij}|$$

Bayes Information Criterion (BIC)

$$BIC(\Lambda) = -\log|\hat{\Omega}(\Lambda)| + tr(S\hat{\Omega}(\Lambda)) + k\frac{log(n)}{n}$$



Multiple precision Matrices: $\Omega = (\Omega^{(1)}, \dots, \Omega^{(K)})$

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \sum_{k=1}^{K} \log|\Omega^{(k)}| - tr(S^{(k)}\Omega^{(k)}) + \mathscr{P}_\lambda(\Omega)$$

Fused graphical LASSO:

$$\mathscr{P}_\lambda(\Omega) = \lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} |\omega_{ij}^{(k)}| + \lambda_2 \sum_{k'>k} \sum_{i,j} |\omega_{ij}^{(k)} - \omega_{ij}^{(k')}|$$

Group graphical LASSO:

$$\mathscr{P}_\lambda(\Omega) = \lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} |\omega_{ij}^{(k)}| + \lambda_2 \sum_{i \neq j} \sqrt{\sum_{k=1}^{K} \omega_{ij}^{(k)^2}}$$

Implementation: R-package(JGL)

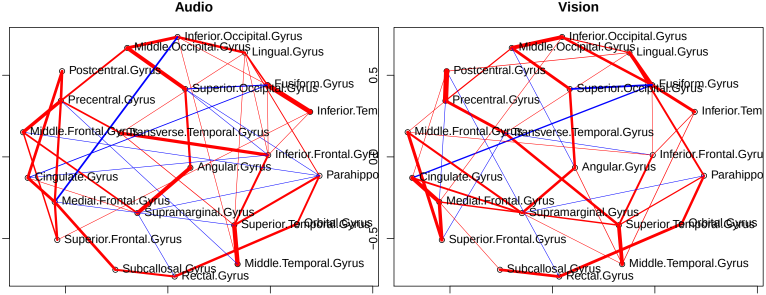
- Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J. & Knight, K.: Sparsity and smoothness via the fused lasso, JRSS B, 2005
- Yang, S.; Lu, Z.; Shen, X.; Wonka, P. & Ye, J.: Fused Multiple Graphical Lasso, see: http://people.math.sfu.ca/ zhaosong
- Danaher, P.; Wang, P. & Witten, D.: The joint graphical lasso for inverse covariance estimation across multiple classes, JRSS B, 2014

Multiple precision Matrices: $\Omega = (\Omega^{(1)}, \ldots, \Omega^{(K)})$

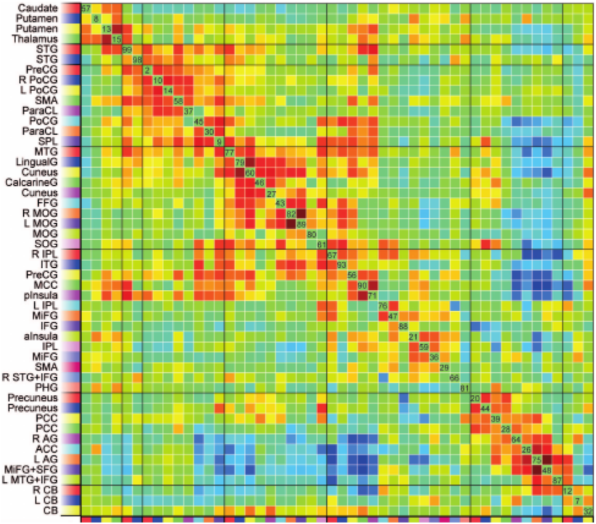
$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmax}} \sum_{k=1}^{K} \log|\Omega^{(k)}| - tr(S^{(k)}\Omega^{(k)}) + \mathscr{P}_\lambda(\Omega)$$

Fused graphical LASSO / SCAD:

$$\mathscr{P}_\lambda(\Omega) = \sum_{k=1}^{K} \sum_{i \neq j} \lambda_{1ij} |\omega_{ij}^{(k)}| + \sum_{k' > k} \sum_{i,j} \lambda_{2ij} |\omega_{ij}^{(k)} - \omega_{ij}^{(k')}|$$

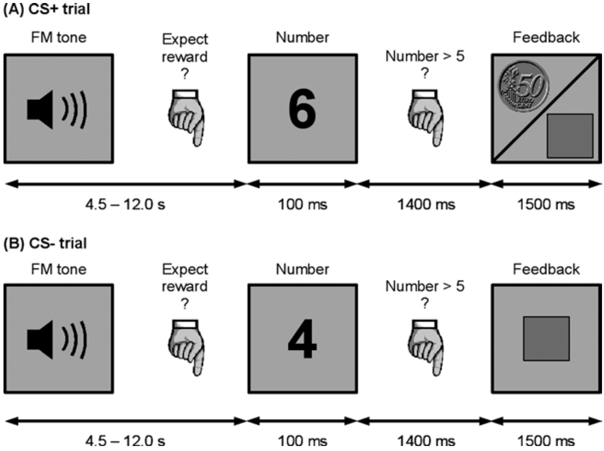


Example: functional connectivity matrix from resting state



Source: Allen et al., Cerebral Cortex 2012.

Learning paradigm:



(Figure: Puschmann (2013))

Interest in changes of brain functionality due to learning

Changes:

- Functional regions becoming active / inactive due to learning
- Changes in sets of regions that act coherently

Classical methods to detect these changes:

- Moving windows or comparison of first third and last third of time series
- Test if parameters / contrasts change over time
- Test if mean value of residuals changes over time
- Test if correlation / partial correlation matrices change over time

Test of stationarity without penalization:

$$H : \Sigma_t \equiv \Sigma \quad \forall t \in (h+1, n-h) :$$

Use (log) Likelihood Ratio Test for

$$H_t : \Sigma_{t-} = \Sigma_{t+}$$

- Can be expressed in terms of eigenvalues $l_1, \ldots, l_p$ of $\hat{\Sigma}_{t-}\hat{\Sigma}_{t+}^{-1}$
- Σ_{t-} and Σ_{t+} estimated from left/right window of size h
- Test-Statistic: $T(l_1, \ldots, l_p) = -C_{h,p} \sum_{i=1}^{p} (\log(l_i) - \log(1 + l_i))$
- $\Rightarrow$ Curves $T(t, h), \quad t \in (h+1, n-h)$
- Distribution under Hypotheses H and H_t does not depend on Σ (as. χ-square)
- Distribution under Hypothesis can be approximated by simulation $\Rightarrow$ density d_h

Problem: Test statistics undefined for $h < p$, highly variable if $h \geq p$

Alternative proposal: Cai and Zhang, Inference for high-dimensional differential correlation matrices. JMVA 2016

Test of stationarity with penalization (GLASSO):

$$H : \Sigma_t \equiv \Sigma \quad \forall t \in (h+1, n-h):$$

Use (log) Likelihood Ratio Test for

$$H_t : \Sigma_{t-} = \Sigma_{t+}$$

- Can be expressed in terms of eigenvalues $l_1, \ldots, l_p$ of $\hat{\Sigma}_{t-}\hat{\Sigma}_{t+}^{-1}$
- Σ_{t-} and Σ_{t+} estimated from left/right window of size h
- Test-Statistic: $T(l_1, \ldots, l_p) = -C_{h,p} \sum_{i=1}^{p} (\log(l_i) - \log(1 + l_i))$
- $\Rightarrow$ Curves $T(t,h), \quad t \in (h+1, n-h)$
- Distribution under Hypotheses H and H_t does depend on Σ

Distribution of test statistic depends on unknown Σ and λ, may be approximated using permutation tests.